



HAL
open science

Modélisation dynamique des sphères anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer : une approche par processus latents

Oladédji Bachirou Tadde

► **To cite this version:**

Oladédji Bachirou Tadde. Modélisation dynamique des sphères anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer : une approche par processus latents. Médecine humaine et pathologie. Université de Bordeaux, 2018. Français. NNT : 2018BORD0306 . tel-02080338

HAL Id: tel-02080338

<https://theses.hal.science/tel-02080338>

Submitted on 26 Mar 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

École doctorale Sociétés, Politique, Santé Publique
Spécialité Santé Publique, option Biostatistique

Par Oladédji Bachirou TADDÉ

MODÉLISATION DYNAMIQUE DES SPHÈRES
ANATOMIQUE, COGNITIVE ET FONCTIONNELLE
DANS LA MALADIE D'ALZHEIMER :
UNE APPROCHE PAR PROCESSUS LATENTS

DYNAMIC MODELING OF ANATOMIC
COGNITIVE AND FUNCTIONAL SPHERES
IN ALZHEIMER'S DISEASE :
A LATENT PROCESS APPROACH

Sous la direction de Cécile PROUST-LIMA

Soutenue publiquement le 30 novembre 2018

Membres du jury

| | | |
|-------------------------|-------------------------|---------------------|
| ANDRIEU Sandrine | PU-PH, INSERM, Toulouse | Présidente |
| LECLERCQ-SAMSON Adeline | PR, CNRS, Grenoble | Rapporteuse |
| NUEL Gregory | PR, CNRS, Paris | Rapporteur |
| JACQMIN-GADDA Hélène | DR, INSERM, Bordeaux | Examinatrice |
| PROUST-LIMA Cécile | CR, INSERM, Bordeaux | Directrice de thèse |

Résumé

En santé publique, l'étude de la progression d'une maladie chronique et de son mécanisme peut nécessiter la modélisation conjointe de plusieurs marqueurs longitudinaux et leur structure de dépendances. Des approches de modélisation existent dans la littérature pour répondre partiellement aux objectifs de modélisation. Ces approches deviennent rapidement coûteuses et difficiles à utiliser dans certaines maladies complexes à caractères latent, dynamique et multidimensionnel.

Dans un tel contexte, l'objectif de cette thèse était de proposer une méthodologie innovante pour modéliser la dynamique de plusieurs processus latents et leurs influences temporelles à partir d'observations répétées de marqueurs continus Gaussiens et non Gaussiens, puis d'appliquer cette approche de modélisation à la maladie d'Alzheimer et les démences apparentées pour modéliser conjointement les dimensions anatomique, cognitive et fonctionnelle atteintes dans la maladie, ainsi que les influences temporelles entre ces dimensions à des fins d'interprétations causales.

L'approche proposée est centrée sur des processus latents qui traduisent les différentes dimensions à modéliser. Elle définit un modèle structurel pour la trajectoire des processus latents et un modèle d'observation pour lier les marqueurs longitudinaux aux processus qu'ils mesurent. Dans le modèle structurel, défini en temps discret, le niveau initial et le taux de variation des processus spécifiques aux individus sont modélisés par des modèles linéaires à effets mixtes. Le modèle du taux de variation comporte une composante autorégressive d'ordre 1 qui permet de modéliser l'effet d'un processus sur un autre processus en tenant explicitement compte du temps. Le modèle structurel tel que défini, bénéficie des mêmes interprétations causales que les modèles à équations différentielles (ODE) de l'approche mécaniste de la causalité tout en évitant les problèmes numériques de ces derniers. Le modèle d'observation utilise des fonctions de lien paramétrées pour que le modèle puisse être appliqué à des marqueurs longitudinaux possiblement non Gaussiens.

Des études de simulations ont validé les bonnes propriétés des estimateurs du maximum de vraisemblance et la justesse de l'inférence de la structures d'influences entre les processus latents malgré une discrétisation faite avec un pas raisonnable. Ces études ont aussi montré que les estimateurs des paramètres du modèle peuvent être biaisés dans le

cas d'une mauvaise spécification de la trajectoire des processus.

Cette approche, appliquée à la modélisation des dimensions anatomique, cognitive, et fonctionnelle dans la maladie d'Alzheimer et des démences apparentées, a permis d'identifier des influences temporelles entre les dimensions anatomique, cognitive et fonctionnelle dans trois stades cliniques de la maladie à partir des données de l'initiative ADNI. Une première application a traité de façon générale chaque dimension puis une seconde application a exploré les relations entre dimensions de façon plus fine. Elle a permis de décrire conjointement la dynamique de l'atrophie de l'hippocampe, du déclin de la mémoire épisodique, du déclin de la fluence verbale et de la perte d'autonomie, ainsi que les influences temporelles entre ces dimensions dans les phases « cognitivement normal », « légers déficits cognitifs » et « diagnostiqué avec une démence de type Alzheimer ».

Mots clés : données répétées, processus latents, causalité, modèles mixtes, équations de différence, maladie d'Alzheimer.

Abstract

In public health, the study of the progression of a chronic disease and its mechanism may require the joint modeling of several longitudinal markers and their dependences structure. Modeling approaches exist in the literature to partially address these modeling objectives. But these approaches are rapidly becoming expensive and difficult to use in some complex diseases involving latent, dynamic and multidimensional aspects.

In this context, the goal of this thesis was to propose an innovative methodology to model the dynamics of several latent processes and their temporal influences from repeated observations of Gaussian and non Gaussian continuous markers, and apply this modeling approach to Alzheimer's disease and related dementias to jointly model the anatomic, cognitive, and functional dimensions of the disease and the temporal influences between these dimensions for causal interpretations. The proposed approach focuses on latent processes that translate the dimensions of interest. It defines a structural model for latent processes and an observation model for the longitudinal process markers. In the structural model, defined in discrete time, the initial level and rate of change of the individual-specific processes are modeled by linear mixed effects models. The model for the rate of change has a first order self-regressive component that is used to model the effect of a process on another process by explicitly accounting for time. The structural model as defined, benefits from the same causal interpretations as the models with differential equations (ODE) in the dynamic causality approach while avoiding numerous problems. The observation model uses parameterized link functions to handle possibly non-Gaussian continuous markers.

Simulation studies have validated the good properties of the maximum likelihood estimators and the accuracy of the inference for the influences structure between the latent processes despite the discretization done with a reasonable step. These simulation studies have also shown that estimators of model parameters may be biased in the case of misspecification of the process trajectory.

This approach, applied to the modeling of anatomic, cognitive, and functional dimensions in Alzheimer's disease, helped identify temporal influences between anatomic, cognitive and functional dimensions in three clinical stages of the disease based on data

from the ADNI data. A first application defined each dimension in a general way. Then, a second application explored more finely the relationships between dimensions. It has allowed to jointly describe the dynamics of hippocampus atrophy, episodic memory decline, decline of verbal fluency, and loss of autonomy as well as temporal influences between these dimensions in cognitively normal subjects, in subjects with mild cognitive impairment and in subjects diagnosed with Alzheimer's dementia.

Key words : longitudinal data, latent processes, causality, mixed models, difference equations, Alzheimer's disease.

Table des matières

| | |
|--|-----------|
| Résumé | 2 |
| Abstract | 4 |
| Remerciements | 7 |
| Production scientifique | 10 |
| Notations et Abréviations | 12 |
| Liste des Tableaux | 16 |
| Liste des Figures | 19 |
| 1 Introduction | 20 |
| 1.1 Maladie d'Alzheimer et démences apparentées | 20 |
| 1.1.1 Définition et contexte épidémiologique | 20 |
| 1.1.2 Signes cliniques et diagnostic | 20 |
| 1.1.3 Facteurs de risque et facteurs protecteurs de la MA | 21 |
| 1.1.4 Modèles théoriques de la maladie d'Alzheimer | 22 |
| 1.1.4.1 Modèle de Jack | 22 |
| 1.1.4.2 Traduction du schéma de Jack en stades | 23 |
| 1.2 Problèmes statistiques liés à la modélisation de la MA | 25 |
| 1.2.1 Modélisation dynamique d'un marqueur répété | 25 |
| 1.2.2 Notion de dimension latente | 26 |
| 1.2.3 Multiplicité des dimensions latentes | 27 |
| 1.2.4 Quelques exemples d'approches pour traiter l'aspect multidimen- sionnel | 27 |
| 1.2.5 Nécessité de nouveaux développements | 29 |
| 1.3 Objectifs de la thèse | 29 |
| 1.3.1 Objectif statistique | 29 |

| | | |
|----------|---|-----------|
| 1.3.2 | Objectif épidémiologique | 29 |
| 1.4 | Données d'étude : l'initiative ADNI | 30 |
| 1.5 | Plan de la thèse | 30 |
| 2 | État de l'art | 31 |
| 2.1 | Modélisation de la dynamique des marqueurs | 31 |
| 2.1.1 | Modélisation de l'évolution d'un marqueur longitudinal | 31 |
| 2.1.1.1 | Modèle linéaire mixte | 31 |
| 2.1.1.2 | Estimation du modèle linéaire mixte | 32 |
| 2.1.1.3 | Robustesse du modèle linéaire mixte aux hypothèses pa- ramétriques et aux données manquantes | 32 |
| 2.1.1.4 | Modélisation de trajectoires non-linéaires | 33 |
| 2.1.1.5 | Extension du modèle linéaire mixte aux données non Gaus- siennes | 33 |
| 2.1.1.6 | Modèle linéaire mixte et notion de processus latent | 34 |
| 2.1.2 | Modélisation conjointe de marqueurs longitudinaux | 35 |
| 2.1.2.1 | Modèle linéaire mixte multivarié | 36 |
| 2.1.2.2 | Prise en compte de la corrélation entre les marqueurs | 37 |
| 2.1.2.3 | Estimation du modèle linéaire mixte multivarié | 38 |
| 2.1.2.4 | Extensions du modèle linéaire mixte multivarié | 38 |
| 2.1.2.5 | Application du modèle linéaire mixte multivarié | 38 |
| 2.1.2.6 | Limites du modèle linéaire mixte multivarié | 39 |
| 2.2 | Modélisation de la causalité | 40 |
| 2.2.1 | Formalismes du problème de causalité | 40 |
| 2.2.1.1 | Schéma causal | 40 |
| 2.2.1.2 | Représentation graphique | 41 |
| 2.2.2 | Approche contrefactuelle de causalité | 44 |
| 2.2.2.1 | Présentation de l'approche contrefactuelle de causalité | 44 |
| 2.2.2.2 | Problème fondamentale de l'approche contrefactuelle de causalité | 45 |
| 2.2.2.3 | Hypothèses et biais dans l'approche contrefactuelle de cau- salité | 45 |
| 2.2.2.4 | Extension de l'approche contrefactuelle de causalité aux études en population | 46 |
| 2.2.2.5 | Modèle marginal structurel | 47 |
| 2.2.3 | Causalité de Granger | 49 |

| | | |
|----------|--|-----------|
| 2.2.3.1 | Présentation de la causalité de Granger | 49 |
| 2.2.3.2 | Régression linéaire multiple | 50 |
| 2.2.3.3 | Réseau Bayésien | 51 |
| 2.2.3.4 | Modèle d'incrémentés linéaires | 55 |
| 2.2.3.5 | Modèle à équations structurelles | 58 |
| 2.2.3.6 | Modèle dynamique à équations structurelles | 60 |
| 2.2.3.7 | Limites de l'approche de causalité de Granger | 62 |
| 2.2.4 | Approche mécaniste de la causalité | 63 |
| 2.2.4.1 | Causalité, notions de filtration et de système autonome | 64 |
| 2.2.4.2 | Décomposition de Doob-Meyer et notion d'orthogonalité | 65 |
| 2.2.4.3 | Modèle statistique à équations différentielles | 66 |
| 2.2.4.4 | Estimation du modèle à équations différentielles | 67 |
| 2.2.4.5 | Applications de l'approche mécaniste de causalité | 68 |
| 2.2.4.6 | Avantages et limites de l'approche mécaniste de causalité | 69 |
| 3 | Méthodologie | 71 |
| 3.1 | Présentation du modèle | 71 |
| 3.1.1 | Modèle structurel | 71 |
| 3.1.1.1 | Présentation du modèle structurel | 71 |
| 3.1.1.2 | Influences temporelles entre les processus | 72 |
| 3.1.1.3 | Corrélations résiduelles entre les processus | 73 |
| 3.1.1.4 | Exemple illustratif du modèle structurel | 74 |
| 3.1.2 | Modèle d'observation pour données complètes | 76 |
| 3.1.2.1 | Modèle d'observation pour marqueur Gaussien | 76 |
| 3.1.2.2 | Modèle d'observation pour marqueur continu non Gaussien | 77 |
| 3.1.2.3 | Écriture matricielle du modèle d'observation | 77 |
| 3.1.3 | Modèle d'observation pour données incomplètes | 79 |
| 3.2 | Estimation du modèle | 79 |
| 3.2.1 | Distributions des processus latents et des observations transformées | 80 |
| 3.2.2 | Identifiabilité du modèle | 81 |
| 3.2.2.1 | Modèle identifiable : définition | 82 |
| 3.2.2.2 | Évaluation de l'identifiabilité du modèle | 82 |
| 3.2.3 | Log-vraisemblance du modèle | 83 |
| 3.2.4 | Maximisation de la log-vraisemblance | 85 |
| 3.2.4.1 | Écriture du problème d'optimisation | 85 |
| 3.2.4.2 | Algorithme d'optimisation | 85 |

| | | |
|----------|---|------------|
| 3.3 | Relations entre matrices d'influences temporelles sous différents pas de discrétisation | 86 |
| 3.4 | Prédictions | 89 |
| 3.4.1 | Prédictions des processus latents | 89 |
| 3.4.2 | Prédictions des observations | 90 |
| 3.4.2.1 | Calcul numérique de l'inverse de la fonction de lien H_k | 90 |
| 3.4.2.2 | Approximation numérique du calcul d'intégrale | 91 |
| 3.5 | Implémentation du modèle sous R | 92 |
| 3.5.1 | Programmes de spécification et de pré-traitement | 92 |
| 3.5.1.1 | Les arguments de spécification du modèle | 93 |
| 3.5.2 | Programmes d'estimation et de prédiction | 95 |
| 3.5.3 | Intervalle de confiance des coefficients de la matrice d'influences temporelles | 96 |
| 3.5.4 | Programmes de sortie des résultats | 97 |
| 4 | Simulations | 98 |
| 4.0.1 | Méthodologie des études de simulation | 98 |
| 4.0.1.1 | Génération des données | 98 |
| 4.0.2 | Validation du programme d'estimation | 99 |
| 4.0.3 | Robustesse du modèle à une mauvaise spécification | 107 |
| 4.0.4 | Mauvaise spécification de modèle et impact de la discrétisation sur la structure d'influences temporelles entre les processus latents | 113 |
| 5 | Applications | 116 |
| 5.1 | Données d'application : ADNI 1 | 116 |
| 5.1.0.1 | Initiative internationale ADNI | 116 |
| 5.1.0.2 | Échantillon d'étude | 117 |
| 5.1.0.3 | Description de l'échantillon | 119 |
| 5.2 | Application aux dimensions anatomique, cognitive et fonctionnelle globales | 122 |
| 5.2.1 | Modélisation | 122 |
| 5.2.1.1 | Modèle | 122 |
| 5.2.1.2 | Estimations des paramètres d'évolution des dimensions | 124 |
| 5.2.1.3 | Estimations des paramètres des modèles d'observation | 127 |
| 5.2.1.4 | Estimations des paramètres de la matrice d'influences temporelles entre les dimensions | 127 |

| | | |
|----------|---|------------|
| 5.2.1.5 | Évaluation de l'adéquation du modèle aux données et analyses de sensibilité | 129 |
| 5.2.2 | Analyse de sensibilité | 129 |
| 5.2.3 | Conclusion | 131 |
| 5.3 | Application 2 : Exploration plus fine des dimensions | 132 |
| 5.3.0.1 | Objectif | 132 |
| 5.3.0.2 | Spécification des modèles | 132 |
| 5.3.1 | Affinement de la dimension anatomique | 133 |
| 5.3.1.1 | Résultats de la modélisation | 134 |
| 5.3.2 | Affinement des dimensions anatomique et cognitive | 138 |
| 5.3.2.1 | Résultats de la modélisation | 139 |
| 5.3.3 | Modélisation des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction » | 142 |
| 5.3.3.1 | Résultats de la modélisation | 145 |
| 5.3.3.2 | Conclusion | 149 |
| 6 | Discussion et Perspectives | 151 |
| 6.1 | Approche méthodologique | 151 |
| 6.1.1 | Objectif | 151 |
| 6.1.2 | Modèle structurel | 151 |
| 6.1.3 | Modèle d'observation | 153 |
| 6.1.4 | Package R développé | 154 |
| 6.1.5 | Étude de simulations | 154 |
| 6.1.6 | Interprétation causale des paramètres de la matrice d'influences temporelles | 154 |
| 6.2 | Épidémiologie de la maladie d'Alzheimer | 156 |
| 6.2.1 | Objectif | 156 |
| 6.2.2 | Application de notre approche à la maladie d'Alzheimer | 156 |
| 6.3 | Perspectives | 158 |
| 6.3.1 | Perspectives de l'approche méthodologique | 158 |
| 6.3.2 | Perspectives des applications | 158 |
| | Bibliographie | 160 |
| | Annexe | 167 |
| 6.1 | Article méthodologique | 167 |
| 6.2 | Complément de résultats de l'application 1 | 203 |

Remerciements

En guise de gratitude pour la chance que la vie m'a accordée et pour l'amour que les Hommes m'ont donné, je voudrais simplement reconduire cette formule simple et fondamentale dans ma vie :

« **Grâce à Dieu, Par les Hommes, Pour les Hommes.** »

Ensuite, le premier Humain que je tiens à remercier pour ce travail de thèse est bien évidemment ma directrice de thèse, Cécile Proust-Lima. Ton enthousiasme, tes connaissances, ton savoir faire, ta pédagogie, ta détermination à avancer malgré les difficultés, ta disponibilité, ta générosité et ton sens de l'écoute sont tant de qualités sans lesquelles ce travail n'aurait jamais pu être fait. Merci infiniment à toi.

Mes remerciements vont aussi à l'endroit des membres du jury pour avoir accepté juger ce travail de thèse. Sandrine Andrieu, je vous remercie d'avoir accepté de présider cette soutenance de thèse. Adeline Leclercq-Samson et Gregory Nuel, je vous remercie d'avoir accepté porter un regard critique sur ce travail entant que rapporteurs. Enfin Hélène Jacqmin-Gadda, je vous remercie d'avoir accepté d'examiner ce travail de thèse.

Je remercie également l'ensemble des équipes de l'ISPED. Notamment les équipes Biostatistique, « Psychoépidémiologie du vieillissement et des maladies chroniques » et SISTEM. Je remercie en particulier Hélène Jacqmin-Gadda, directrice de l'équipe Biostatistique pour m'avoir accueilli dans l'équipe, Daniel Commenges pour l'initiation à la causalité, Jean-François Dartigues pour ses idées et sa détermination à trouver des applications des modèles biostatistiques, Viviane Philipps pour ta collaboration, Loïc Ferrer (le Sètois) pour les informations et supports.

Je remercie également le Mésocentre de Calcul Intensif Aquitain (MCIA) pour m'avoir donné accès au cluster Avakas sans lequel les résultats d'analyse de cette thèses ne seraient pas aussi enrichis.

REMERCIEMENT

Je remercie le professeur Vincent Couallier pour m'avoir enseigné les statistiques et pour m'avoir parlé de ce stage à l'INSERM encadré par Cécile Proust-Lima : ce fut le début de l'histoire.

Je remercie tous mes collègues de bureaux Student et 40, Laura Vilain, Camille Sabathé, Corentin Ségalas, Mathilde Reydit de m'avoir supporté.

Un merci du fond du coeur à Joëlle Courreges, Anaïs Rouanet, Boris Hejblum et Diénabou Sylla, pour leur soutien dans les moments difficiles et leur sens de l'écoute.

Mes derniers remerciements vont à l'endroit de mes amis et de ma famille. Je pense en particulier à Raphaël Bekoin, tu es comme un père et à mon meilleur ami Ibrahim Dahoui (one two three viva l'algérie).

Je remercie mon oncle Ismaël Taddé pour avoir contribué à faire de moi l'homme que je suis devenu. Enfin je conclus mes remerciements en pensant à mon feu père Taddé Aboudou, à ma mère restée au pays (loin des yeux mais près du coeur), à mes frères et soeurs, mon épouse Khèfilath Amoussa et mes deux filles Maïdath et Maïssa.

Ce travail a bénéficié de l'appui financier de l'Agence Nationale de la Recherche dans le cadre du projet « Modèles Statistiques pour la Maladie d'Alzheimer et le vieillissement » (SMALA) (ANR-15-CE37-0002).

REMERCIEMENT

*Je dédie ce manuscrit à mes filles, Maïdath
et Maïssa.*

Production scientifique

Article

Publication

- ▶ Taddé O. B., Jacqmin-Gadda H., Dartigues J-F., Commenges D. et Proust-Lima C. Dynamic modeling of multivariate latent processes and their temporal relationships : application to Alzheimer's disease. *Biometrics, révision majeure*
- ▶ Taddé O. B., Dartigues J-F. et Proust-Lima C. Temporal relationships between cerebral atrophy, cognitive functioning and functional dependency in prodromal Alzheimer's disease. *en préparation*
- ▶ Carles S., Taddé O. B., et al. Dynamic Reciprocal relationships between cognitive and functional declines along the Alzheimer's Dementia continuum *en préparation*

Communications

Communications orales

- ▶ Taddé O.B., et Proust-Lima C. Modélisation dynamique de la causalité entre processus latents : application aux sphères anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer. 48^{èmes} *Journées de Statistique de la Société Française de Statistique*, Montpellier, France, 2016.
- ▶ Taddé O. B. et Proust-Lima C. Modelling causal links between anatomic, cognitive and functional dimensions in context of Alzheimer's disease : a latent process approach. *Journée de l'école doctorale SP2*, Bordeaux, France, 2017.
- ▶ Taddé O. B., Jacqmin-Gadda H., Commenges D. et Proust-Lima C. Modelling trajectories and causal links between dimensions in Alzheimer's Disease : a latent process approach. *International Society for Clinical Biostatistics Conference*, Vigo, Espagne, 2017.

PRODUCTION SCIENTIFIQUE

► Taddé O. B., Jacqmin-Gadda H., Dartigues J-F., Commenges D. et Proust-Lima C. Dynamic modeling of latent processes and their causal relationships : application to Alzheimer's disease. *Journée des jeunes chercheurs de la société française de biométrie*, Paris, France, 2018.

Communication écrite (poster)

► Taddé O. B., Jacqmin-Gadda H., Commenges D. et Proust-Lima C. Modelling trajectories and causal links between dimensions in Alzheimer's Disease : a latent process approach. *International Biometric Society Conference*, Barcelone, Espagne, 2018.

Notations et Abréviation

Notations

- Λ : Un processus latent
- δ : Pas de discrétisation
- $\Delta\Lambda(t + \delta) = \Lambda(t + \delta) - \Lambda(t)$: Incrément de Λ entre les instants t et $t + \delta$
- $\frac{d\Lambda(t)}{dt}$: Dérivée de $\Lambda(t)$ par rapport au temps
- Y : Marqueur d'intérêt
- X : Variable explicative, variable cause
- \mathbf{Q} : en gras signifie que Q est un vecteur, ou une matrice
- $\mathbb{E}[Y]$: Espérance de la variable aléatoire Y
- $\mathbb{E}[Y|X]$: Espérance de la variable aléatoire Y conditionnellement à la variable X
- $\mathbb{P}[A]$: Probabilité de A
- $\mathbb{P}[A|B]$: Probabilité de A conditionnellement à B
- $\mathcal{N}(\mu, \Sigma)$: Distribution normale (possiblement multivariée) d'espérance μ et de variance Σ
- $\mathcal{L}(\cdot)$: Fonction de vraisemblance
- $L(\cdot)$: Fonction de log-vraisemblance
- \mathbb{I}_D : Matrice identité de taille $D \times D$
- $\boldsymbol{\theta}$: Vecteur de paramètres
- $\hat{\boldsymbol{\theta}}$: Estimation de $\boldsymbol{\theta}$
- u, v : Effet aléatoire
- \mathbf{B} : Matrice de variance covariance des effets aléatoires
- \mathbf{A} : Matrice d'influences temporelles
- \mathbf{L} : Matrice inférieure de la décomposition de Cholesky.

Abréviations

| | | |
|-------------|---|--|
| LDG | : | Local Dependency Graph |
| DAG | : | Directed Acyclic Graph |
| MSM | : | Marginal Structural Model |
| SEM | : | Structural Equation Modeling |
| ODE | : | Ordinary Differential Equations |
| EM | : | Expectation-Maximisation algorithm |
| SAEM | : | Stochastic Approximation of Expectation-Maximisation Algorithm |
| AIC | : | Akaike Information Criterion |
| MA | : | Maladie d'Alzheimer |
| AD | : | Alzheimer's dementia |
| MCI | : | Mild Cognitive Impairment |
| CN | : | Cognitively Normal |

Liste des tableaux

| | | |
|-----|---|-----|
| 3.1 | Liste des sous-arguments de l'argument <code>Structural.model</code> du package R <code>CInLPN</code> | 94 |
| 3.2 | Liste des sous-arguments de l'argument <code>measurement.model</code> du package R <code>CInLPN</code> | 95 |
| 4.1 | Étude de validation par simulation du programme d'estimation (scénario 1) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture à 95% de 96 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1. | 101 |
| 4.2 | Étude de validation par simulation du programme d'estimation en présence de données manquantes (scénario 2) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 96 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1 avec environ 25% de taux de visites manquantes et une probabilité de 0.05 qu'au moins un marqueur soit manquant à une visite effectuée. | 103 |
| 4.3 | Étude de validation par simulation du programme d'estimation pour un modèle présentant une structure d'influences temporelles entre les processus qui dépend du temps (scénario 3 à gauche sans données manquantes, scénario 4 à droite avec données manquantes) : Estimation moyenne ($\hat{\theta}$), écart-type empirique (ETE), écart-type asymptotique (ETA) et taux de couverture à 95% (TC) de 200 répliques d'un échantillon de 512 sujets. | 106 |
| 4.4 | Résultats des simulations en cas de mauvaise spécification des trajectoires des processus (scénario 5) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets. | 108 |

NOTATIONS ET ABRÉVIATIONS

| | | |
|-----|--|-----|
| 4.5 | Résultats des simulations en cas de mauvaise spécification des trajectoires des processus (scénario 6) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets. | 110 |
| 4.6 | Résultats des simulations en cas d'omission des corrélations entre les niveaux initiaux des processus latents (scénario 7) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1 ; estimation des paramètres du modèles sans la prise en compte des corrélations résiduelles au temps $t = 0$ entre les niveaux des processus latents. | 111 |
| 4.7 | Impact de la discrétisation sur la structure d'influences temporelles entre les processus : Erreur de type I (en %) associée à chaque paramètre d'influence $a_{dd'}$ avec $(d, d') = \{1, 2, 3\}^2$ et $d \neq d'$ mesurant l'influence du processus d' sur le procesus d pour des modèles estimés avec les pas de discrétisation $\delta = 1/3, 1/2, 1$ sur un échantillon simulé en temps quasi continu avec un solveur ($\delta = 0.001$). L'intervalle de confiance à 95% de l'erreur de type -I attendue est de $[1.98, 8.02]$ pour l'erreur nominale de 5%. | 114 |
| 5.1 | Statistiques descriptives de l'échantillon de 656 sujets de la base de données ADNI 1 répartis en trois groupes de sujets : CN, MCI et AD. | 119 |
| 5.2 | Estimations des effets fixes des modèles du niveau initial et du taux de variation des 3 dimensions pour un pas de discrétisation $\delta = 0.23 \approx 3$ mois. | 125 |
| 5.3 | Estimations des paramètres de la matrice d'influences temporelles reliant les taux de variation sur un intervalle d'amplitude $\delta = 0.23 \approx 3$ mois des dimensions anatomique, cognitive et fonctionnelle à l'état précédent de ces dimensions. | 128 |
| 5.4 | Tableau comparatif des estimations (valeur) des éléments de la matrice d'influences temporelles et de leur p-valeur (p) pour cinq spécifications de modèles. | 131 |
| 5.5 | Estimations des paramètres fixes du modèle conjoint des dimensions « atrophie de l'hippocampe », « cognition globale » et « fonction ». | 134 |
| 5.6 | Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe » et « cognition ». | 140 |

| | | |
|-----|--|-----|
| 5.7 | Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction ». | 145 |
| 6.1 | Estimations des paramètres non contraints de la décomposition de Cholesky (\mathbf{L}) de la matrice de variance-covariance $\mathbf{B} = \mathbf{L}^\top \mathbf{L}$ des effets aléatoires $(u_{i1}, u_{i2}, u_{i3}, v_{i1}^1, v_{i1}^2, v_{i2}^1, v_{i2}^2, v_{i3}^1, v_{i3}^2)^\top$ du modèle multivarié de la dimension anatomique mesurée par le composite score de 10 régions du cerveau (Anat10), de la dimension cognitive mesurée par le composite score de 19 scores psychométriques standardisés (Cog19) et la dimension fonctionnelle mesurée par la somme standardisée de 30 items de la FAQ (Fct); les taux de variation des dimensions sont supposés linéaires au cours du temps (modèle de l'application 1 de la section 5.2.1.1). | 203 |
| 6.2 | Estimations des écart-types des erreurs de mesure σ_{Anat10} du marqueur Anat10 (composite score de 10 régions du cerveau qui mesure la dimension anatomique), σ_{Cog19} du marqueur Cog19 (composite score de 19 scores psychométriques standardisés qui mesure la dimension cognitive), σ_{Fct} du marqueur Fct (somme standardisée de 30 items de la FAQ qui mesure la dimension fonctionnelle) et des paramètres des fonctions de lien par I-splines intégrées quadratiques avec 4 noeuds pour les marqueurs Anat10 , Cog19 et 3 noeuds pour le marqueur Fct (cf. modèle principal de l'application 1 de la section 5.2.1.1). | 204 |

Table des figures

| | | |
|-----|---|----|
| 1.1 | Schéma théorique de la maladie d'Alzheimer impliquant les dimensions biologique, cérébrale, cognitive et fonctionnelle [Jack et al., 2010]. | 23 |
| 1.2 | Schéma hypothétique de l'évolution séquentielle du processus pathologique de la maladie d'Alzheimer conduisant à l'altération de la dimension cognitive [Sperling et al., 2011] | 24 |
| 1.3 | Schéma hypothétique des relations entre les dimensions cognitive, fonctionnelle et comportementale [McLaughlin et al., 2010] | 25 |
| 2.1 | DAG représentant un système Λ de trois composantes Λ_1 , Λ_2 et Λ_3 observées une seule fois. | 41 |
| 2.2 | DAG représentant un système Λ de trois composantes Λ_1 , Λ_2 , Λ_3 observées de manière répétée dans le temps. Les instances du système au temps t peuvent influencer les instances du système au temps $t + 1$ | 42 |
| 2.3 | Graphe de dépendance locale représentant un système Λ de trois composantes Λ_1 , Λ_2 , Λ_3 ; dans le LDG, les composantes du système Λ sont stochastiques et un arc orienté d'une composante $\Lambda_{d'}$ vers une composante Λ_d indique l'effet de la composante $\Lambda_{d'}$ sur le changement instantané de la composante Λ_d | 43 |
| 2.4 | DAG correspondant au schéma causal modélisé dans un modèle de régression linéaire multiple où Y_1 est la variable dépendante, Z une variable de confusion et Y_2 la variable indépendante. | 51 |
| 2.5 | DAG représentant un modèle à équations structurelles pour un système Λ de trois composantes latentes Λ_1 , Λ_2 et Λ_3 mesurées respectivement par les variables Y_1 , Y_2 et Y_3 . α_{21} , α_{31} et α_{32} modélisent les effets causaux entre les composantes du système Λ | 59 |

| | | |
|-----|--|-----|
| 3.1 | Exemple d'une base de I-splines et d'une fonction de lien. La figure de gauche représente une base de splines intégrées de degré 2 avec 3 noeuds aux points 0, 5 et 12 utilisée pour transformer un marqueur à valeurs dans $[0,24]$. La figure de droite représente la fonction de lien associée avec les paramètres de régression $(-4.70, \exp(1.80), \exp(-1.67), \exp(-2.06), \exp(2.48), \exp(-1.18))$ | 78 |
| 5.1 | Histogrammes des marqueurs dans les échelles naturelle et transformée, et fonctions de lien estimées pour les dimensions anatomique, cognitive et fonctionnelle). | 122 |
| 5.2 | Trajectoires moyennes prédites avec intervalles de confiance des dimensions anatomique, cognitive, fonctionnelle selon le stade clinique (CN, MCI, AD) pour deux profils de sujets : profil p1 = femme porteuse de l'allèle apolipoprotéine $\epsilon 4$ ayant fait moins de 12 ans d'étude; profil p2 = femme non porteuse de l'allèle apolipoprotéine $\epsilon 4$ ayant fait plus de 12 ans d'étude. . . | 126 |
| 5.3 | Structure d'influences temporelles entre les dimensions anatomique (Λ_A), cognitive (Λ_C) et fonctionnelle (Λ_F) aux trois stades cliniques de la maladie d'Alzheimer (CN, MCI, AD). Une flèche de Λ_{d1} vers Λ_{d2} indique l'influence temporelle de la dimension $d1$ au temps t sur le taux de variation de la dimension Λ_{d2} au temps $t + \delta$ après ajustement sur le niveau des autres dimensions au temps t et sur les variables explicatives des trajectoires des dimensions; δ étant le pas de discrétisation. | 127 |
| 5.4 | Comparaison dans chaque stade clinique de la moyenne des observations transformées à chaque temps d'observation discret à la moyenne des prédictions. | 129 |
| 5.5 | Schéma de modélisation des dimensions « atrophie de l'hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l'hippocampe (Hip), « cognition globale » (Λ_C) mesurée par un composite score de 19 scores cognitifs standardisés (Cog19), dimension fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (Fct). | 134 |
| 5.6 | Influences temporelles entre les dimensions « atrophie de l'hippocampe » (Λ_A), « cognitive globale » (Λ_C) et fonctionnelle (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer). | 138 |

LISTE DES FIGURES

5.7 Schéma de modélisation des dimensions « atrophie de l'hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l'hippocampe (**Hip**), « cognition » (Λ_C) mesurée à la fois par le composite score de sous-scores standardisés du test « Auditory Verbal Learning Test » (**Avlt**) et par le score standardisé au test « Category Fluency Test » dans le champ lexical animal (**CatFlu**), fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (**Fct**). 139

5.8 Influences temporelles entre les dimensions « atrophie de l'hippocampe » (Λ_A), « cognition » (Λ_C) et fonction (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer). 140

5.9 Schéma de modélisation des dimensions « atrophie de l'hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l'hippocampe (**Hip**), la dimension « mémoire épisodique » (Λ_M) mesurée par un composite score de sous-scores standardisés du tests « Auditory Verbal Learning Test » (**Avlt**), la dimension « fluence verbale » (Λ_L) mesurée par le score standardisé au test « Category Fluency Test » pour la catégorie animale (**CatFlu**) et la dimension fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (**Fct**). 143

5.10 Influences temporelles entre les dimensions « atrophie de l'hippocampe » (Λ_A), « mémoire épisodique » (Λ_M), « fluence verbale » (Λ_L) et « fonction » (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer). 149

1 Introduction

1.1 Maladie d'Alzheimer et démences apparentées

1.1.1 Définition et contexte épidémiologique

La maladie d'Alzheimer (MA) est une maladie neuro-dégénérative qui se développe sur des dizaines d'années avant le diagnostic clinique. Elle conduit progressivement à la perte d'autonomie et à la démence. Elle est la principale cause de démence et concerne principalement les personnes âgées de plus de 65 ans.

En 2015, la maladie d'Alzheimer et les démences apparentées ont touché plus de 27 millions de personnes, et d'ici 2050 plus de 106 millions de personnes seront touchées par la maladie d'Alzheimer [Brookmeyer et al., 2007]. Dans les pays où l'espérance de vie est en augmentation comme la France, l'incidence de la maladie d'Alzheimer et des démences en général est plus importante et entraîne un coût important de prise en charge des personnes touchées par la maladie.

1.1.2 Signes cliniques et diagnostic

La maladie d'Alzheimer est une maladie multidimensionnelle caractérisée par les dysfonctionnements progressifs de plusieurs fonctions notamment biologiques, cérébrales, et cognitives. Elle est un processus continu de vieillissement pathologique avec une phase asymptomatique et une phase symptomatique.

La phase asymptomatique est caractérisée par les atteintes de la dimension biologique notamment le dysfonctionnement de la protéine β -amyloïde, et les atteintes de la dimension cérébrale caractérisées par le dysfonctionnement de la protéine Tau, la perte de synapses et la dégénérescence des neurones qui forment des enchevêtrements neurofibrillaires [Weiner et al., 2012].

Pendant la phase symptomatique, les dysfonctionnements de la phase asymptomatique s'accroissent et les fonctions cognitives commencent à être atteintes avec une altération notamment de la mémoire, de l'attention, du langage et des fonctions exécutives. Ces

INTRODUCTION

dysfonctionnements progressent jusqu'à la perte d'autonomie du sujet [Amieva et al., 2008].

Jusqu'à aujourd'hui, le diagnostic de la maladie d'Alzheimer et plus généralement de la démence est essentiellement posé dans la phase symptomatique sur la base des altérations de la mémoire et d'au moins une autre fonction cognitive objectivées par des tests psychométriques, et de la perte d'autonomie dans les activités de la vie quotidienne objectivées par des échelles de dépendance. Cependant le diagnostic définitif de la maladie d'Alzheimer ne peut être donné que post-mortem après autopsie du cerveau et l'observation des lésions cérébrales (plaques séniles et dégénérescences neuro-fibrillaires) [Reitz et Mayeux, 2014].

Les récents travaux pointant les dimensions atteintes dans la phase asymptomatique ont conduit cette année à une redéfinition des critères de diagnostic de la maladie d'Alzheimer. Il s'agit des critères AT(N) [Jack et al., 2018]. Suivant les critères AT(N), le diagnostic de la maladie d'Alzheimer et des démences apparentées devrait à l'avenir être posé à partir des dysfonctionnements des bio-marqueurs de la protéine β -Amyloïde, de la protéine **T**au et de la **N**eurodégénérescence objectivée par des examens d'imagerie médicale.

L'ensemble de ces critères diagnostiques témoignent clairement de l'aspect multidimensionnel de la maladie d'Alzheimer avec l'implication des dimensions biologique et cérébrale très tôt dans la phase asymptomatique puis des dimensions cognitive et fonctionnelle dans la phase symptomatique. Une vision globale de la dynamique de ces dimensions et de leurs relations est cruciale pour d'une part mieux comprendre l'histoire naturelle de la maladie d'Alzheimer, son étiologie et ses facteurs de risque, et d'autre part pour évaluer de nouveaux traitements en développement.

1.1.3 Facteurs de risque et facteurs protecteurs de la MA

Les facteurs de risque et les facteurs protecteurs potentiels de la maladie d'Alzheimer et plus généralement des démences sont d'ordre génétique, médical, comportemental et socio-démographique.

Parmi les facteurs socio-démographiques, le vieillissement (le grand âge) est le premier facteur de risque de la maladie d'Alzheimer. L'incidence de la maladie d'Alzheimer et des démences apparentées augmente exponentiellement avec l'âge jusqu'à la tranche d'âge de 80-90 ans puis l'effet de l'âge s'atténue après 90 ans [Jorm et Jolley, 1998; Gao et al.,

1998].

L'incidence de la démence est plus élevée chez les femmes que chez les hommes [Letenneur et al., 1999]. Toutefois, Joly et al. [2002] à l'aide d'un modèle multi-état prenant en compte le décès et la censure par intervalle ont montré que la différence d'incidence de la démence entre les hommes et les femmes est spécifique à l'âge. L'incidence de la démence est plus élevée chez les hommes que chez les femmes avant 80 ans puis après 80 ans l'incidence est plus élevée chez les femmes que chez les hommes. La hausse de l'incidence chez les femmes par rapport aux hommes s'expliquerait par le risque de décès plus élevé chez les hommes après 80 ans.

Un faible niveau d'éducation est un facteur de risque de la démence en général et de la maladie d'Alzheimer en particulier [Letenneur et al., 1999; Jorm et Jolley, 1998].

Les facteurs médicaux associés à la maladie d'Alzheimer sont les antécédents de traumatisme crânien qui augmentent le risque de la maladie d'Alzheimer chez l'homme mais pas chez la femme [Fleminger et al., 2003]; et les antécédents de dépression Jorm [2001]. Toutefois la dépression pourrait aussi être perçue comme un symptôme de la phase symptomatique pré-démentielle de la maladie d'Alzheimer.

La consommation de cigarettes augmenterait aussi légèrement le risque de la maladie [Almeida et al., 2002].

Le niveau élevé d'éducation à travers la réserve cérébrale, la consommation modérée d'alcool et la prise de médicaments anti-inflammatoires sont des facteurs protecteurs de la maladie d'Alzheimer [Wickett et al., 2000; Jorm et al., 1997; Broe et al., 2000; Orgogozo et al., 1997].

1.1.4 Modèles théoriques de la maladie d'Alzheimer

1.1.4.1 Modèle de Jack

Les aspects dynamique et multidimensionnel de la maladie d'Alzheimer ont été traduits sous forme de schémas théoriques (hypothétiques) pour appréhender la maladie d'Alzheimer dans son ensemble. Le schéma de Jack et al. [2010], reproduit en Figure 1.1, conceptualise la succession de dysfonctionnements asymptomatiques avec principalement une accumulation de protéines dans le cerveau (β -amyloïde, Tau), une atrophie cérébrale (notamment de l'hippocampe) suivis de manifestations cliniques caractérisées par une altération des fonctions cognitives qui conduisent à la perte d'autonomie.

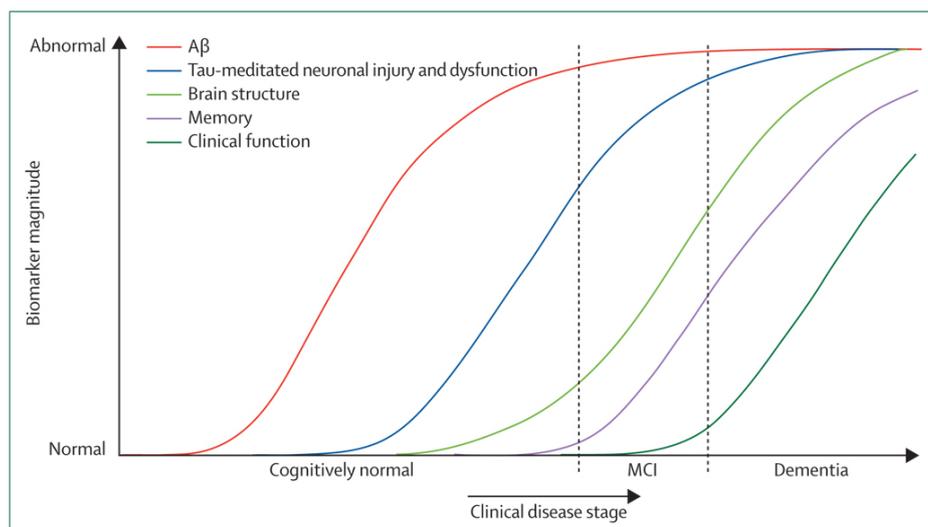


FIGURE 1.1 – Schéma théorique de la maladie d'Alzheimer impliquant les dimensions biologique, cérébrale, cognitive et fonctionnelle [Jack et al., 2010].

Suivant le schéma publié en 2010 puis mis à jour en 2013 [Jack et al., 2010, 2013], les auteurs ont défini trois étapes importantes dans le développement de la maladie d'Alzheimer et des démences apparentées. Il s'agit de l'étape CN (Cognitively Normal) où les dysfonctionnements ne sont pas suffisamment importants et le sujet ne présente pas de déficits cognitifs différenciables du processus de vieillissement cognitif normal, l'étape MCI (Mild Cognitive Impairment) qui est caractérisée par de légers déficits cognitifs et l'étape AD (Alzheimer's dementia) où le sujet est diagnostiqué dément.

Le schéma théorique de Jack [Jack et al., 2010, 2013] est aujourd'hui devenu central dans la recherche sur la maladie d'Alzheimer.

1.1.4.2 Traduction du schéma de Jack en stades

Sperling et al. [2011] ont traduit le modèle de Jack sous forme d'un schéma reliant de manière séquentielle les atteintes des dimensions biologique, cérébrale et cognitive (schéma reproduit en Figure 1.2). La maladie est décrite en trois étapes correspondant à l'accumulation successive des atteintes à ces trois dimensions.

En se focalisant sur la phase symptomatique de la maladie, McLaughlin et al. [2010] ont proposé un schéma hypothétique qui relie les atteintes des dimensions cognitive, fonctionnelle, et comportementale entre elles en amont et en aval du diagnostic (schéma reproduit en Figure 1.3).

Enfin, cette année, Jack et al. [2018] ont proposé une description plus fine la maladie

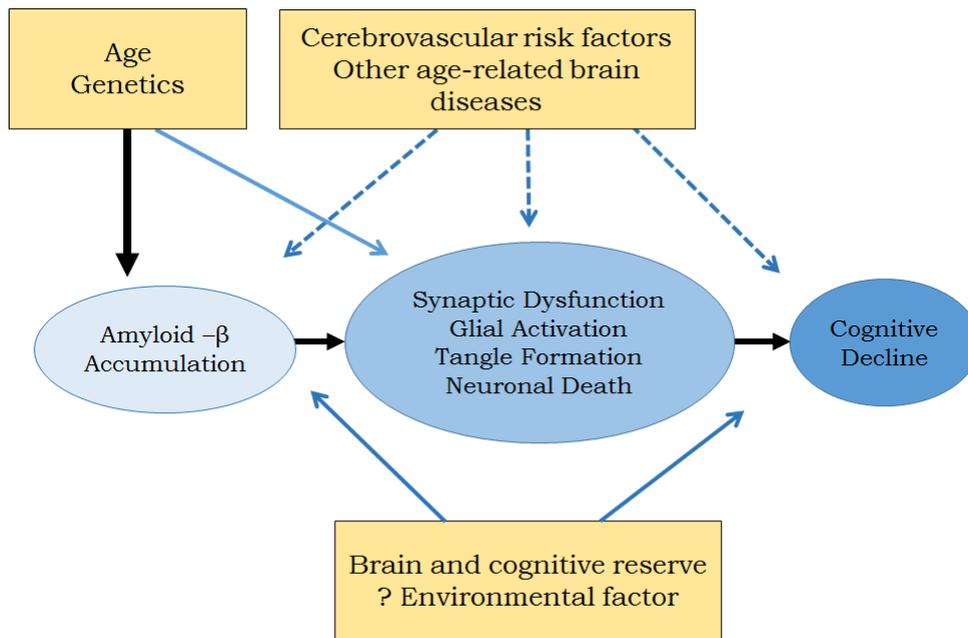


FIGURE 1.2 – Schéma hypothétique de l'évolution séquentielle du processus pathologique de la maladie d'Alzheimer conduisant à l'altération de la dimension cognitive [Sperling et al., 2011]

d'Alzheimer en six étapes que sont :

- **Étape 1** : Elle est la première étape de la phase asymptomatique qui débute par les dysfonctionnements de la protéine β -amyloïde entraînant la formation de plaques amyloïdes.
- **Étape 2** : Elle est la deuxième étape de la phase asymptomatique. Cette étape est caractérisée par la poursuite de la formation des plaques amyloïdes à laquelle s'ajoutent le dysfonctionnement de la protéine Tau, la perte de synapses et la dégénérescence des neurones qui forment des enchevêtrements neuro-fibrillaires. Les enchevêtrements neuro-fibrillaires conduisent à l'atrophie cérébrale de plusieurs régions du cerveau, notamment de l'hippocampe qui est central dans le processus de mémorisation.
- **Étape 3** : Elle est la première étape de la phase symptomatique. À cette étape, les dépôts de plaques amyloïdes et les enchevêtrements neuro-fibrillaires se poursuivent et les premiers signes d'altération cognitive apparaissent. Il s'agit de légers troubles cognitifs qui ne sont pas suffisamment importants pour différencier le processus pathologique de vieillissement cognitif du processus normal de vieillissement cognitif.

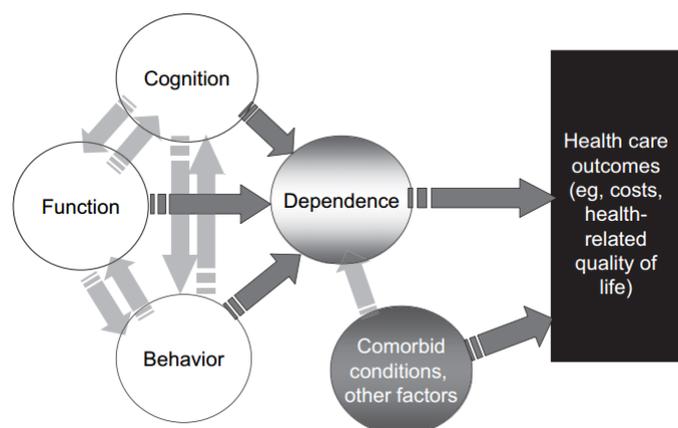


FIGURE 1.3 – Schéma hypothétique des relations entre les dimensions cognitive, fonctionnelle et comportementale [McLaughlin et al., 2010]

- **Étape 4** : Elle est qualifiée de MCI (en anglais, Mild Cognitive Impairment). Cette étape est principalement caractérisée par l’accentuation du déclin cognitif conduisant à de légères plaintes cognitives objectivées par des tests psychométriques.
- **Étape 5** : Elle correspond à la première étape de la phase démentielle. Cette étape est caractérisée par le diagnostic d’une démence modérée (en anglais moderate dementia).
- **Étape 6** : Cette étape est caractérisée par un état de démence sévère.

Ces schémas ont le mérite de proposer une conceptualisation large de la maladie d’Alzheimer qui tient compte de ses aspects dynamique, multidimensionnel, et des relations entre les dimensions impliquées dans la maladie. Néanmoins, ils restent la plupart du temps des postulats, alimentés par diverses connaissances théoriques et observations sur des données non nécessairement longitudinales. Ils doivent donc maintenant être traduits en modèles statistiques estimés sur de grandes bases de données de population afin d’être confirmés ou révisités.

1.2 Problèmes statistiques liés à la modélisation de la Maladie d’Alzheimer

1.2.1 Modélisation dynamique d’un marqueur répété

La maladie d’Alzheimer et les démences apparentées sont des pathologies dynamiques qui se développent sur une grande période de temps. Pour étudier leur progression, des

marqueurs de la maladie, comme des tests psychométriques et des examens d'imagerie médicale par exemple, sont collectés de façon répétée au cours du temps dans des études de cohorte.

Pour comprendre l'aspect dynamique de la maladie d'Alzheimer et des démences apparentées, il est important de modéliser ces marqueurs répétés de la maladie. Le modèle linéaire à effets mixtes est souvent utilisé pour modéliser l'évolution d'un marqueur et l'effet de variables explicatives sur le marqueur [Laird et Ware, 1982]. Il est adapté pour des marqueurs Gaussiens. Cependant les marqueurs de la maladie d'Alzheimer peuvent être de différents types, discret (catégoriel, ordinal), continu Gaussien, ou continu avec une distribution asymétrique dont il faut tenir compte dans la modélisation [Proust-Lima et al., 2011].

1.2.2 Notion de dimension latente

Une autre difficulté liée à la modélisation de la maladie d'Alzheimer est celle de la multiplicité des marqueurs répétés de la maladie. Pour mesurer avec plus d'acuité une même quantité appelée dimension latente (ou trait latent), plusieurs marqueurs sont simultanément collectés. Par exemple, le processus de vieillissement cognitif est une notion abstraite non observable directement mais qui peut être mesurée par une batterie de tests psychométriques répétés dans le temps. Ces tests mesurent différentes facettes de la cognition. Certains vont mesurer un fonctionnement cognitif global, c'est le cas du Mini-Mental State Examination (MMSE), test extrêmement connu qui évalue sous forme d'items diverses dimensions cognitives (e.g., mémoire, langage, représentation dans l'espace). D'autres vont être plus spécifiques. Par exemple, certains mesurent la fluence verbale via la capacité à citer des mots dans un certain champ lexical, c'est le cas du Set Test d'Isaacs (IST). D'autres vont mesurer les fonctions exécutives comme le Trail Making Test (TMT) qui consiste à relier des lettres dans l'ordre alphabétique (version A) et alternativement des lettres et des chiffres croissants (version B).

Pour bien comprendre la dynamique d'une dimension latente (par exemple le déclin cognitif), il est important de modéliser conjointement les divers marqueurs répétés de cette dimension afin d'en capter toutes ses spécificités. L'extension multivariée du modèle linéaire mixte peut être utilisée pour modéliser conjointement la dynamique de plusieurs marqueurs, l'effet de variables explicatives sur ces marqueurs ainsi que leurs corrélations.

Cependant, cette approche ne permet pas de modéliser directement la dynamique d'une dimension latente et les effets de variables explicatives sur cette dimension à travers les observations de plusieurs marqueurs de la dimension. Une alternative est le modèle

linéaire mixte à processus latent qui permet de modéliser conjointement plusieurs marqueurs d'une même dimension ainsi que la dynamique de la dimension latente. L'idée ici est de supposer que tous les marqueurs mesurent un même processus latent dont l'évolution est décrite par un modèle mixte, et de spécifier une équation d'observation pour chaque marqueur qui le relie au processus latent [Proust-Lima et al., 2006] dans l'esprit des modèles à équations structurelles [Duncan, 1975].

1.2.3 Multiplicité des dimensions latentes

La maladie d'Alzheimer et les démences apparentées sont des pathologies multidimensionnelles avec l'implication entre autres d'une dimension anatomique à travers l'atrophie cérébrale, d'une dimension cognitive à travers le déclin de fonctions cognitives telles que la mémoire ou l'attention et d'une dimension fonctionnelle se traduisant par la perte progressive de l'autonomie. Ces dimensions sont potentiellement latentes, dynamiques et reliées entre elles.

La compréhension du mécanisme de développement de la maladie d'Alzheimer et plus généralement des démences, ainsi que la compréhension de la structure des liens entre les dimensions impliquées peut nécessiter des développements statistiques. Il est en effet nécessaire de modéliser conjointement des marqueurs de plusieurs dimensions, la dynamique des dimensions au cours du temps, la structure des relations entre les dimensions et l'effet de variables explicatives.

1.2.4 Quelques exemples d'approches pour traiter l'aspect multidimensionnel

Des approches de modélisation ont été proposées dans la littérature pour tenir compte des différents problèmes liés à la modélisation de la maladie d'Alzheimer et des démences apparentées. Nous revenons ici sur deux exemples pour traiter en particulier la multiplicité des dimensions dynamiques.

Le modèle linéaire mixte multivarié peut être utilisé pour modéliser conjointement plusieurs marqueurs possiblement de différentes dimensions. Par exemple, pour étudier spécifiquement des sous fonctions cognitives, Robitaille et al. [2012] ont utilisé le modèle linéaire mixte bivarié et modélisé conjointement les évolutions d'un test de performances visuelles (test visuo-spatial de Kohs [Dureman et Sälde, 1959]) et d'un test de fonctions exécutives (le Digit-Symbol Substitution Test [Wechsler, 1991]). Ils ont ainsi pu décrire

l'évolution des deux marqueurs et leurs facteurs de risque respectifs en prenant en compte la corrélation possible entre les marqueurs. Cette approche reste cependant descriptive. D'une part, elle ne permet pas de travailler sur les dimensions latentes sous-jacentes. D'autre part, elle ne permet pas d'appréhender les relations entre les marqueurs au delà de corrélations entre les effets aléatoires décrivant la dynamique de chaque marqueur. Il s'en suit que cette approche est difficile d'interprétations causales car ce sont les marqueurs et leurs corrélations qui sont modélisés et non directement les relations temporelles entre dimensions d'intérêts.

Une autre approche consiste à modéliser les relations temporelles entre des marqueurs de plusieurs dimensions en ajustant systématiquement la dynamique d'un marqueur d'une dimension sur des caractéristiques temporelles d'autres dimensions (par exemple, le niveau initial, la pente ou le niveau courant) qui sont supposées être observées sans erreurs de mesure et indépendantes [Landau et al., 2011; Han et al., 2012]. Par exemple, Landau et al. [2011] ont utilisé cette approche pour examiner chez des sujets aux stades MCI et AD, les relations entre le métabolisme du glucose dans les régions angulaires, post cingulaires et temporales du cerveau mesuré par le marqueur FDG-PET (Fludeoxyglucose-Positron Emission Tomography), le déclin cognitif mesuré par les scores du test ADAS-cog (Alzheimer Disease Assessment Scale), un test évaluant le niveau cognitif global, et les atteintes fonctionnelles mesurées par la FAQ (Functional Assessment Questionnaire).

Dans ces modèles, les relations directionnelles entre dimensions sont directement appréhendées en se focalisant plus sur des corrélations entre les marqueurs. Cependant, tous les marqueurs ne sont pas considérés comme stochastiques et ces modèles tendent à faire l'hypothèse (possiblement forte) d'une séquence prédéterminée des atteintes des dimensions.

Les deux approches mentionnées ci-dessus ne sont pas satisfaisantes pour étudier les relations entre les dimensions de progression de la maladie d'Alzheimer et des démences. D'une part dans ces approches, ce sont les marqueurs des dimensions qui sont directement modélisés et non les dimensions latentes elles-même si bien que les conclusions de ces modèles sont relatives aux marqueurs et ne sont pas extrapolables facilement dans l'échelle des dimensions. D'autre part, soit les relations entre les dimensions ne sont pas explicitement modélisées (dans le modèle mixte multivarié) soit l'aspect stochastique des dimensions n'est pas proprement pris en compte (dans les régressions d'une dimension sur d'autres).

1.2.5 Nécessité de nouveaux développements

Des développements statistiques sont donc nécessaires pour fournir une approche de modélisation simultanée des multiples dimensions de la maladie d'Alzheimer qui tient compte de leur aspect latent, de leur aspect dynamique et des relations temporelles entre les dimensions. La finalité de ces nouveaux développements est d'offrir une approche globale de modélisation de la maladie dont la prise en compte des relations entre les dimensions contribuera à réduire les biais d'estimation sur les paramètres des dimensions et à appréhender les mécanismes causaux entre les dimensions.

Du point de vue épidémiologique, la perspective est une meilleure compréhension de l'histoire naturelle de la maladie et des mécanismes qui lient les dimensions entre elles. Spécifiquement, un objectif est de confronter les différents schémas théoriques aux données réelles. A long terme, une approche dynamique et multidimensionnelle vise à fournir une meilleure prédiction de l'évolution de la maladie d'Alzheimer et des démences apparentées pour poser précocement leur diagnostic et identifier plus en amont des facteurs de risque de la maladie d'Alzheimer et des démences apparentées pour la mise en place de politiques préventives.

1.3 Objectifs de la thèse

1.3.1 Objectif statistique

L'objectif de la thèse est de proposer une nouvelle approche statistique permettant de modéliser conjointement les dynamiques de plusieurs processus latents et de capter leurs interactions temporelles. Plutôt que d'explorer des corrélations descriptives entre les dimensions, nous modéliserons directement les relations temporelles entre les dimensions afin de bénéficier d'interprétations causales sous certaines hypothèses.

Nous emprunterons pour cela d'une part la méthodologie des modèles mixtes pour décrire l'évolution des dimensions latentes, définies comme des processus latents mesurés par plusieurs marqueurs répétés. D'autre part, nous utiliserons la méthode des modèles mécanistes pour décrire le lien entre la dynamique des processus latents.

1.3.2 Objectif épidémiologique

L'objectif épidémiologique de la thèse est d'appliquer notre approche statistique pour décrire conjointement les trajectoires de différentes dimensions impliquées dans la maladie

d'Alzheimer, particulièrement les dimensions anatomique, cognitive et fonctionnelle et leurs relations causales.

1.4 Données d'étude : l'initiative ADNI

Les méthodes et outils développés dans cette thèse seront appliqués aux données de la cohorte internationale ADNI.

L'initiative ADNI (Alzheimer's Disease Neuroimaging Initiative) est une étude multidisciplinaire qui a pour objectif de collecter sur le long terme des bio-marqueurs génétiques, cliniques, biochimiques et d'imagerie (IRM) à des fins d'études relatives à la maladie d'Alzheimer. L'initiative ADNI a été conçue pour trouver des méthodes plus sensibles et précises de détection précoce de la maladie d'Alzheimer, de suivi de sa progression et d'évaluation des effets de traitements. L'étude, lancée en octobre 2004, incluait initialement 400 sujets atteints de troubles cognitifs légers, 200 sujets diagnostiqués en phase initiale de la maladie d'Alzheimer et 200 sujets témoins (sujets normaux). Actuellement dans sa troisième phase (ADNI2, après ADNI GO, ADNI 1), l'étude constitue une riche source de données à la base de nombreuses publications internationales [Weiner et al., 2012].

1.5 Plan de la thèse

La suite de la thèse est organisée autour de quatre chapitres. Le chapitre 2 présente l'état de l'art des méthodes statistiques utiles aux travaux de la thèse. Le chapitre 3 décrit l'approche de modélisation proposée. Dans le chapitre 4 nous présentons les résultats d'études de simulations visant à valider notre méthodologie. Le chapitre 5 présente une série d'applications aux données des cohortes ADNI. Enfin le chapitre 6 est consacré à la discussion et aux conclusions de la thèse.

2 État de l'art

Ce chapitre présente un état de l'art des méthodes statistiques pour modéliser l'évolution de marqueurs longitudinaux, leurs associations et leurs relations causales. Ce chapitre est scindé en deux sections. La première section est consacrée à la théorie des modèles mixtes pour décrire la trajectoire de marqueurs au cours du temps. Dans la deuxième section nous aborderons les approches de modélisation causale.

2.1 Modélisation de la dynamique des marqueurs

2.1.1 Modélisation de l'évolution d'un marqueur longitudinal

2.1.1.1 Modèle linéaire mixte

Considérons n sujets indicés $i = 1, \dots, n$ ayant chacun n_i mesures répétées du marqueur Y , $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})$, aux temps de mesure t_{ij} , $j = 1, \dots, n_i$. Le modèle linéaire à effets mixtes introduit par Laird et Ware [1982] est très utilisé pour modéliser la trajectoire d'un marqueur longitudinal. Il modélise les effets fixes linéaires de variables explicatives sur l'évolution moyenne du marqueur tout en prenant en compte la corrélation entre les mesures répétées d'un même sujet au travers d'effets aléatoires propres au sujet. L'écriture matricielle du modèle linéaire à effets mixtes pour le vecteur \mathbf{Y}_i s'écrit comme suit :

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i \quad (2.1)$$

où \mathbf{X}_i est une matrice de taille $n_i \times p$ contenant p variables explicatives aux n_i répétitions du marqueur. La matrice \mathbf{X}_i peut inclure des variables explicatives indépendantes ou dépendantes du temps. $\boldsymbol{\beta}$ est le vecteur d'effets fixes associés aux p variables explicatives. \mathbf{Z}_i est une sous-matrice de \mathbf{X}_i de taille $n_i \times q$ associée aux effets aléatoires \mathbf{u}_i propres à chaque sujet i . Les effets aléatoires sont Gaussiens centrés de matrice de variance-covariance \mathbf{B} : $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$. Le vecteur d'erreurs de mesure $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{in_i})$ est supposé Gaussien centré de matrice de variance-covariance $\boldsymbol{\Sigma}_i$: $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$. Les erreurs ϵ_{ij} contenues dans le vecteur $\boldsymbol{\epsilon}_i$ sont mutuellement indépendantes, et indépendantes des effets aléatoires \mathbf{u}_i : $\boldsymbol{\Sigma}_i = \sigma^2\mathbb{I}_{n_i}$.

Les effets aléatoires servent à capter la variabilité entre les sujets et à capturer la corrélation entre les mesures répétées d'un même sujet. Afin de prendre en compte de manière plus fine la corrélation entre les mesures répétées d'un même sujet i , le vecteur d'erreurs de mesure $\boldsymbol{\epsilon}_i$ peut être décomposé en des erreurs stochastiques \boldsymbol{w}_i corrélées et des erreurs de mesures \boldsymbol{e}_i indépendantes. La matrice $\boldsymbol{\Sigma}_i$ devient $\boldsymbol{R}_i + \sigma^2 \mathbb{I}_{n_i}$, où \boldsymbol{R}_i et $\sigma^2 \mathbb{I}_{n_i}$ sont respectivement les matrices de variance-covariance des erreurs stochastiques \boldsymbol{w}_i et des erreurs de mesures \boldsymbol{e}_i . Un exemple d'erreurs stochastiques utilisé est le mouvement Brownien de fonction de covariance $\text{cov}(w_i(t_{ij}), w_i(t_{ij'})) = \omega^2 \min(t_{ij}, t_{ij'})$: $\boldsymbol{R}_i = (\omega^2 \min(t_{ij}, t_{ij'}))_{j,j' \in \{1, \dots, n_i\}^2}$.

2.1.1.2 Estimation du modèle linéaire mixte

La formulation marginale du modèle linéaire mixte (2.1) s'écrit :

$$\boldsymbol{Y}_i \sim \mathcal{N}(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{V}_i = \boldsymbol{Z}_i \boldsymbol{B} \boldsymbol{Z}_i^\top + \boldsymbol{\Sigma}_i). \quad (2.2)$$

À partir de la formulation marginale du modèle linéaire mixte, la vraisemblance du modèle est donnée par :

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n \left(\frac{1}{2\pi} \right)^{n_i/2} |\boldsymbol{V}_i|^{-1/2} \exp \left(-\frac{1}{2} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta})^\top \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{X}_i \boldsymbol{\beta}) \right) \quad (2.3)$$

où $\boldsymbol{\theta}$ est le vecteur de paramètres contenant le vecteur des effets fixes $\boldsymbol{\beta}$, les paramètres des matrices \boldsymbol{B} et $\boldsymbol{\Sigma}_i$. L'estimateur du maximum de vraisemblance du vecteur $\boldsymbol{\theta}$ est obtenu par maximisation itérative du logarithme népérien de la vraisemblance (2.3). La matrice de variance-covariance des estimations est obtenue en inversant la matrice d'information de Fisher.

Les algorithmes de type Newton-Raphson (exemple, l'algorithme de Marquardt [1963]) et les algorithmes EM [Dempster et al., 1977] sont utilisés pour l'estimation du vecteur de paramètres $\boldsymbol{\theta}$ [Lindstrom et Bates, 1988; Verbeke, 1997]. L'approche Bayésienne associée à l'échantillonneur de Gibbs est aussi utilisée pour l'estimation des paramètres du modèle [Lange et al., 1992].

2.1.1.3 Robustesse du modèle linéaire mixte aux hypothèses paramétriques et aux données manquantes

Le modèle linéaire mixte a pour objectif principal de corriger les biais d'estimation sur les effets fixes dans le modèle linéaire classique, en tenant compte des corrélations entre les mesures répétées des sujets. Ces corrélations sont captées par les effets aléatoires et/ou les erreurs stochastiques corrélées.

La robustesse de l'estimation des paramètres des effets fixes à des écarts aux hypothèses sur les distributions des effets aléatoires et des erreurs de mesures a été investiguée dans des travaux de simulation. Jacqmin-Gadda et al. [2006] et Warrington et al. [2014] ont par exemple montré que les estimateurs des effets fixes étaient robustes dans le cas d'erreurs de mesure non Gaussiennes ou hétéroscédastiques mais n'étaient pas robustes dans le cas d'erreurs de mesures corrélées entre elles. Verbeke et Lesaffre [1997] ont montré que les estimateurs des effets fixes et des composantes de variance étaient non-biaisés et asymptotiquement Gaussiens dans le cas d'effets aléatoires non Gaussiens. Toutefois, une correction de l'inverse de la matrice d'information de Fisher est nécessaire pour obtenir les estimations asymptotiques non-biaisées des écart-types du vecteur de paramètres θ .

L'estimateur du maximum de vraisemblance est robuste aux données manquantes aléatoires [Laird, 1988], c'est-à-dire au processus de données manquantes qui peut être prédit par les observations.

2.1.1.4 Modélisation de trajectoires non-linéaires

Le modèle linéaire mixte est linéaire par rapport aux effets fixes et aléatoires mais il n'est pas forcément limité à la modélisation de trajectoires linéaires dans le temps. Il permet de décrire diverses formes de trajectoires du marqueur selon les fonctions du temps incluses dans la matrice \mathbf{X}_i .

Dans les premiers modèles en VIH, des fonctions linéaires du temps ont été utilisées pour décrire l'évolution du marqueur CD4 et de la charge virale [Lange et al., 1992]. Depuis, des trajectoires plus flexibles sont appréhendées. En vieillissement cognitif, des fonctions quadratiques et plus généralement polynomiales du temps sont souvent utilisées pour tenir compte de l'accélération du déclin cognitif dans le grand âge [Wilkosz et al., 2010].

Pour modéliser des trajectoires plus complexes on peut aussi utiliser des bases de splines ou des polynômes fractionnaires [Verbyla et al., 1999; Green et Silverman, 1993].

2.1.1.5 Extension du modèle linéaire mixte aux données non Gaussiennes

Le modèle linéaire mixte est restreint aux données continues Gaussiennes alors qu'en pratique le marqueur d'intérêt peut être non Gaussien.

Le modèle linéaire a été généralisé aux données de distribution de la famille exponentielle (données binaires, catégorielles ou de comptage) [Zeger et Liang, 1986; Dunson, 2003]. La distribution f_{Y_i} des observations Y_{ij} du marqueur Y est de la forme :

$$f_{Y_i}(y_{ij}; \theta_{ij}, \psi) = \exp\left(\psi^{-1}(\theta_{ij}y_{ij} - g(\theta_{ij})) + C(y_{ij}, \psi)\right) \quad (2.4)$$

où la fonction dérivée g' de g établit le lien entre le moment d'ordre 1 du marqueur Y et le paramètre canonique $\boldsymbol{\theta}_{ij} : \mathbb{E}[Y_{ij}] = g'(\boldsymbol{\theta}_{ij})$.

Pour capturer la corrélation entre les mesures répétées des sujets et la variabilité inter-sujet, le paramètre canonique $\boldsymbol{\theta}_{ij}$ est modélisé par un modèle linéaire mixte :

$$\boldsymbol{\theta}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_i, \quad (2.5)$$

où $\boldsymbol{\beta}$ sont les effets fixes associés aux variables explicatives \mathbf{X}_{ij} et \mathbf{u}_i sont les effets aléatoires Gaussiens associés aux variables explicatives \mathbf{Z}_{ij} .

Notons que le modèle linéaire généralisé inclut le modèle linéaire mixte standard pour une fonction de lien g' égale à la fonction identité : $\mathbb{E}[Y_{ij}|u_i] = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_i$.

Comme alternative au modèle linéaire généralisé, Proust-Lima et al. [2006, 2013] ont proposé un modèle qui adjoint au modèle linéaire mixte une transformation paramétrée du marqueur Y_{ij} pour se ramener à un marqueur intermédiaire Gaussien \tilde{Y}_{ij} . Le modèle s'écrit :

$$\begin{cases} \tilde{Y}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}\mathbf{u}_i + \epsilon_{ij} \\ Y_{ij} = H^{-1}(\tilde{Y}_{ij}, \boldsymbol{\eta}), \end{cases} \quad (2.6)$$

où H est une fonction monotone de vecteur de paramètres $\boldsymbol{\eta}$. Tous les paramètres du modèle sont estimés simultanément en maximisant le logarithme népérien de la vraisemblance.

Les fonctions H sont choisies dans des familles souples de transformations (Beta CDF standardisée, Weibull CDF standardisée, combinaison de base de I-splines) à l'aide du critère d'Akaike [Akaike, 1973]. Si le modèle de transformation n'est pas contraint, des contraintes d'identifiabilité doivent être ajoutées au modèle linéaire mixte du marqueur transformé \tilde{Y} pour que le modèle soit globalement identifiable.

Notons que le modèle linéaire mixte est un cas particulier de cette approche dans lequel la fonction de transformation H est la fonction identité ou une fonction linéaire des paramètres de transformation.

2.1.1.6 Modèle linéaire mixte et notion de processus latent

Généralement, les mesures répétées d'un marqueur longitudinal sont relatives à un phénomène d'intérêt qui peut être défini statistiquement par un processus latent. En reprenant l'écriture du modèle linéaire mixte plus général tel que défini en (2.6) on peut faire ressortir explicitement le processus sous-jacent continu $(\boldsymbol{\Lambda}_t)_{t \geq 0}$ qui est mesuré indirectement par le marqueur Y_{ij} à des temps discrets $(t_{ij})_{j=1, \dots, n_i}$ chez l'individu $i = 1, \dots, n$, soit :

$$\begin{cases} \Lambda_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{Z}_{it}\mathbf{u}_i \\ \tilde{Y}_{ij} = \Lambda_{it_{ij}} + \epsilon_{ij} \\ Y_{ij} = H^{-1}(\tilde{Y}_{ij}, \boldsymbol{\eta}), \end{cases} \quad (2.7)$$

où $(\boldsymbol{\Lambda}_t)_{t \geq 0}$ est le processus latent continu sous-jacent au marqueur d'intérêt Y_{ij} dont la transformée \tilde{Y}_{ij} par la fonction H est l'observation bruitée du processus latent avec une erreur de mesure ϵ_{ij} supposée Gaussienne. Tous les paramètres du modèle sont estimés simultanément de la même manière que dans le modèle linéaire à transformation monotone de Proust-Lima et al. [2006, 2013].

L'avantage de l'écriture du modèle (2.7) est qu'il permet de dissocier le sous-modèle structurel pour le processus latent du sous-modèle d'observation pour le marqueur. Cette approche de modélisation offre d'une part la possibilité de varier le modèle d'observation suivant la nature des observations pour un même modèle structurel. D'autre part, elle permet d'assouplir explicitement le modèle structurel en rajoutant un processus d'erreurs stochastiques $(w_t)_{t \geq 0}$ pour mieux modéliser la dynamique du processus latent [Diggle, 1988].

Par exemple, dans la maladie d'Alzheimer un processus latent d'intérêt peut être le niveau cognitif. Ce processus n'est pas observable directement. Mais il peut être mesuré à l'aide de plusieurs tests psychométriques qui constituent des marqueurs longitudinaux du processus cognitif [Proust-Lima et al., 2007].

Cette approche de modélisation distinguant le modèle structurel du modèle d'observation est fréquente en psychométrie. Il s'agit des modèles à traits latents, ou des modèles IRT (Item Response Theory) lorsque les items sont catégoriels. Dans ces modèles, à chaque temps d'observation t la quantité abstraite d'intérêt (aptitude, faculté, intelligence, etc) est modélisée par une variable latente θ_t et non pas par un processus latent $(\boldsymbol{\theta}_t)_{t \geq 0}$; la variable θ_t est ensuite reliée à un ou plusieurs items par une fonction de lien. Ces modèles servent souvent à évaluer les qualités métrologiques de différents items mesurant une même quantité abstraite [Baker, 2001].

2.1.2 Modélisation conjointe de marqueurs longitudinaux

Le modèle linéaire mixte est classiquement utilisé pour modéliser les mesures répétées d'un marqueur longitudinal ou d'une dimension grâce à l'introduction d'un processus latent. Cependant, dans les études longitudinales, il est fréquent de collecter plusieurs marqueurs de plusieurs dimensions possiblement corrélés entre eux. La modélisation conjointe

de ces marqueurs permet d'une part d'analyser simultanément leurs évolutions et leur structure de dépendance, et d'autre part d'utiliser cette structure de dépendance et l'ensemble des informations disponibles pour augmenter la puissance des analyses. Dans cette section nous allons présenter essentiellement la version multivariée du modèle linéaire mixte.

2.1.2.1 Modèle linéaire mixte multivarié

Pour modéliser conjointement les dynamiques de plusieurs marqueurs longitudinaux et leurs corrélations, la théorie du modèle linéaire mixte à été étendue au cadre multivarié.

Présentation du modèle linéaire mixte multivarié

Soient $\mathbf{Y}_1, \dots, \mathbf{Y}_K$ K marqueurs longitudinaux. Pour un sujet $i = 1, \dots, n$ $\mathbf{Y}_{ik} = (Y_{i1k}, \dots, Y_{ijk}, \dots, Y_{in_{ik}k})$ est le vecteur de taille n_{ik} des observations du $k^{\text{ième}}$ marqueur aux temps d'observation $(t_{ijk})_{j=0, \dots, n_{ik}}$. Le modèle linéaire mixte avec une composante stochastique \mathbf{w}_{ik} pour le $k^{\text{ième}}$ marqueur s'écrit :

$$\mathbf{Y}_{ik} = \mathbf{X}_{ik}\boldsymbol{\beta}_k + \mathbf{Z}_{ik}\mathbf{u}_{ik} + \boldsymbol{\epsilon}_{ik} \quad (2.8)$$

où $\mathbf{X}_{ik}, \boldsymbol{\beta}_k, \mathbf{Z}_{ik}, \mathbf{u}_{ik}, \boldsymbol{\epsilon}_{ik} = \mathbf{w}_{ik} + \mathbf{e}_{ik}$ sont spécifiques au marqueur k et sont définis comme dans la section 2.1.1.1 : $\mathbf{u}_{ik} \sim \mathcal{N}(0, \mathbf{B}_k)$ et $\boldsymbol{\epsilon}_{ik} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{ik})$, avec $\boldsymbol{\Sigma}_{ik} = \mathbf{R}_{ik} + \sigma_k^2 \mathbb{I}_{n_{ik}}$.

En posant $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iK})$; $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K)$, $\mathbf{u}_i = (\mathbf{u}_{i1}, \dots, \mathbf{u}_{iK})$, $\mathbf{w}_i = (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iK})$, $\mathbf{e}_i = (e_{i1}, \dots, e_{iK})$, l'écriture matricielle du modèle linéaire mixte (2.8) s'écrit :

$$\mathbf{Y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{u}_i + \boldsymbol{\epsilon}_i \quad (2.9)$$

où \mathbf{X}_i et \mathbf{Z}_i sont des matrices bloc-diagonales de $k^{\text{ème}}$ élément respectif \mathbf{X}_{ik} et \mathbf{Z}_{ik} . Le vecteur des effets aléatoires \mathbf{u}_i est Gaussien centré de matrice de variance-covariance \mathbf{B} et indépendant du vecteur d'erreurs $\boldsymbol{\epsilon}_i = \mathbf{w}_i + \mathbf{e}_i$ qui est aussi un vecteur Gaussien centré de matrice de variance-covariance $\boldsymbol{\Sigma}_i = \mathbf{R}_i + \mathbf{M}_i$: $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$ et $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$.

Les erreurs stochastiques \mathbf{w}_i sont des réalisations d'un processus stochastique Gaussien multivarié de matrice de variance-covariance \mathbf{R}_i et le vecteur d'erreurs de mesures indépendantes \mathbf{e}_i est Gaussien de matrice de variance \mathbf{M}_i , une matrice bloc diagonale de bloc k qui vaut $\sigma_k^2 \mathbb{I}_{n_{ik}}$.

De manière analogue au modèle linéaire mixte univarié défini en (2.7), le modèle linéaire mixte multivarié peut être formulé de manière à faire ressortir un processus latent multivarié $\Lambda_{t(t \geq 0)}$ pour lequel on spécifie un modèle structurel puis un modèle d'observation pour relier les marqueurs répétés $\mathbf{Y}_i = (\mathbf{Y}_{i1}, \dots, \mathbf{Y}_{iK})$ au processus latent $\Lambda_{t(t \geq 0)}$.

2.1.2.2 Prise en compte de la corrélation entre les marqueurs

Pour des raisons de clarté, nous détaillerons la modélisation de la corrélation entre marqueurs dans le modèle linéaire mixte bivarié ($K = 2$). Cependant la procédure est la même pour un nombre plus important de marqueurs. La corrélation entre les marqueurs peut être modélisée à l'aide de la matrice de variance-covariance des effets aléatoires et/ou de la matrice de variance-covariance des erreurs stochastiques. Dans le cas du modèle linéaire mixte bivarié, la matrice de variance-covariance des effets aléatoires \mathbf{u}_i , des erreurs stochastiques \mathbf{w}_i et des erreurs de mesures \mathbf{e}_i s'écrivent :

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_{12} \\ \mathbf{B}_{12}^\top & \mathbf{B}_2 \end{bmatrix}, \quad \mathbf{R} = \begin{bmatrix} R_1 & R_{12} \\ R_{12}^\top & R_2 \end{bmatrix}, \quad \mathbf{M}_i = \begin{bmatrix} \sigma_1^2 \mathbb{I}_{n_{i1}} & \mathbf{0} \\ \mathbf{0} & \sigma_2^2 \mathbb{I}_{n_{i2}} \end{bmatrix}$$

Les matrices de variance-covariance \mathbf{B}_1 et \mathbf{B}_2 des effets aléatoires intra-marqueurs \mathbf{u}_{i1} et \mathbf{u}_{i2} et les variances R_1 et R_2 des erreurs stochastiques intra-marqueurs w_{i1} et w_{i2} servent généralement à modéliser les corrélations entre les mesures répétées intra-marqueurs. Elles sont spécifiques au marqueur et permettent de modéliser au mieux la dynamique intrinsèque de chaque marqueur en tenant compte des corrélations entre les mesures répétées de chaque sujet.

La matrice \mathbf{B}_{12} de covariances des effets aléatoires inter-marqueurs ($\mathbf{u}_{i1}, \mathbf{u}_{i2}$) et la covariance R_{12} des erreurs stochastiques w_{i1} et w_{i2} servent à modéliser les possibles corrélations entre les mesures répétées du marqueur Y_1 et du marqueur Y_2 . De manière spécifique, la matrice \mathbf{B}_{12} capte la part des corrélations inter-marqueurs liée à la variabilité inter-sujet et la covariance R_{12} capte la corrélation entre les mesures répétées des deux marqueurs liée aux erreurs stochastiques.

Fieuws et Verbeke [2006] ont par exemple modélisé conjointement plusieurs marqueurs longitudinaux en se servant de la matrice de variance-covariance des effets aléatoires sur les intercepts et les pentes des marqueurs pour capter les corrélations inter-marqueurs. Sy et al. [1997] ont modélisé la corrélation entre deux marqueurs longitudinaux à l'aide de la matrice de variance-covariance du processus bivarié IOU (Intégré Ornstein-Uhlenbeck) utilisé pour modéliser les erreurs stochastiques.

2.1.2.3 Estimation du modèle linéaire mixte multivarié

La vraisemblance du modèle linéaire mixte multivarié s'écrit de manière analogue à celle d'un modèle linéaire mixte univarié donnée en (2.3). Les mêmes procédures de maximisation du logarithme népérien de la vraisemblance et les mêmes algorithmes d'optimisation peuvent être utilisés pour estimer les paramètres du modèle. Thiébaud et al. [2002] ont proposé une procédure d'estimation d'un modèle linéaire mixte bivarié de deux marqueurs (Y_1, Y_2) à partir de la procédure Proc MIXED de SAS d'estimation d'un modèle linéaire mixte univarié. L'idée de Thiébaud et al. [2002] est de créer un seul marqueur Y qui met bout-à-bout les observations des deux marqueurs en leur associant leur temps de mesure respectif et un vecteur binaire indiquant le marqueur mesuré à chaque temps. Cette approche peut être facilement adoptée sous d'autres logiciels (exemple R).

Des problèmes de calcul numérique surviennent dès que le nombre de marqueurs et/ou le nombre d'effets aléatoires utilisés par marqueur augmentent. Fieuws et Verbeke [2006] ont proposé une approche d'estimation basée sur les pseudo-vraisemblances de tous les paires possibles des K marqueurs prises séparément pour en déduire les estimations de tous les paramètres du modèle complet qui inclut tous les K marqueurs.

2.1.2.4 Extensions du modèle linéaire mixte multivarié

Toutes les extensions du modèle linéaire mixte présentées en univarié (modèle linéaire mixte généralisé, modèle avec fonction de lien paramétrée, modèle à processus latent) sont applicables au cadre multivarié. Les mêmes procédures d'estimation que dans le cas univarié sont utilisées.

2.1.2.5 Application du modèle linéaire mixte multivarié

Le modèle linéaire mixte multivarié trouve de nombreuses applications en épidémiologie notamment dans les études sur le VIH pour étudier simultanément les trajectoires de charge virale et de CD4 [Sy et al., 1997; Thiébaud et al., 2002] ou bien dans le vieillissement cognitif [Robitaille et al., 2012; Mungas et al., 2005; Hall et al., 2001]. Par exemple, comme évoqué en introduction, dans le processus normal de vieillissement cognitif Robitaille et al. [2012] ont utilisé un modèle linéaire mixte bivarié pour modéliser conjointement le déclin de la vitesse d'exécution mesurée par le score répété au Digit-Symbol Substitution Test (noté Y_{ij1}) [Wechsler, 1991] et le déclin de la capacité visuo-spatiale mesurée par le score répété au Kohs Block Design Test (noté Y_{ij2}) [Kohs, 1923]. La corrélation entre les deux marqueurs est captée par des effets aléatoires corrélés (spécifiques au individu et

au marqueurs) sur le niveau initial et la pente des deux marqueurs. Leur modèle s'écrit :

$$Y_{ijk} = \beta_{0k} + \beta_{1k}\text{age}0_i + \beta_{2k}\text{sexe}_i + u_{i0k} + \beta_{3k}t_{ijk} + u_{i1k}t_{ijk} + \epsilon_{ijk}, \quad k = 1, 2 \quad (2.10)$$

avec Y_{ijk} l'observation du marqueur $k = 1, 2$ chez le sujet i à la visite j , $\beta_{.k}$ les effets fixes du modèle, $\text{age}0_i$ l'âge du sujet i à l'inclusion, age_{ijk} l'âge du sujet i à la visite j , $t_{ijk} = \text{age}_{ijk} - \text{age}0_i$, ϵ_{ijk} des erreurs de mesure Gaussiennes centrées de variance σ_k^2 et indépendantes : $\epsilon_{ij1} \perp\!\!\!\perp \epsilon_{ij2}$, u_{i0k} et u_{i1k} sont respectivement des effets aléatoires sur l'intercept et la pente du marqueur k avec

$$\left((u_{i0k}, u_{i1k})_{k=1,2}^\top \right) \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_{u_{01}}^2 & \sigma_{u_{01}u_{11}} & \sigma_{u_{01}u_{02}} & \sigma_{u_{01}u_{12}} \\ \sigma_{u_{11}u_{01}} & \sigma_{u_{11}}^2 & \sigma_{u_{11}u_{02}} & \sigma_{u_{11}u_{12}} \\ \sigma_{u_{02}u_{01}} & \sigma_{u_{02}u_{11}} & \sigma_{u_{02}}^2 & \sigma_{u_{02}u_{12}} \\ \sigma_{u_{12}u_{01}} & \sigma_{u_{12}u_{11}} & \sigma_{u_{12}u_{02}} & \sigma_{u_{12}}^2 \end{bmatrix} \right)$$

Sur les données OCTO-Twin [Johansson et McClearn, 1996], Robitaille et al. [2012] ont trouvé une association entre l'évolution de la vitesse d'exécution et l'évolution de la capacité visuo-spatiale qui se traduit par des corrélations significatives intercept-intercept et pente-pente entre les effets aléatoires des marqueurs des deux dimensions latentes.

2.1.2.6 Limites du modèle linéaire mixte multivarié

Le modèle linéaire à effets mixtes est utilisé pour modéliser conjointement les dynamiques de plusieurs marqueurs répétés et pour explorer les corrélations entre les marqueurs. Cette approche de modélisation permet de réduire les biais d'estimation sur les paramètres de chaque marqueur et d'améliorer les prédictions des marqueurs.

La principale limite de cette approche est la difficulté d'interprétations causales des paramètres d'association entre les marqueurs répétés. Lorsque l'objectif est de comprendre les mécanismes causaux entre des phénomènes dynamiques, cette approche atteint ses limites. En effet, les corrélations entre les marqueurs sont des associations bi-directionnelles qui n'informent pas sur le sens des dépendances entre les marqueurs. Par exemple, si on considère deux marqueurs répétés Y_1 et Y_2 générés par un mécanisme causal dans lequel le marqueur Y_1 affecte le marqueur Y_2 , la modélisation conjointe des deux marqueurs par un modèle linéaire mixte bivarié peut permettre d'identifier une corrélation entre les deux marqueurs dont la prise en compte va permettre d'améliorer l'estimation des paramètres de chaque marqueur. Mais le modèle linéaire mixte bivarié n'informe pas plus sur le sens de la relation causale entre Y_1 et Y_2 .

2.2 Modélisation de la causalité

La modélisation causale a pour but d'inférer des relations causales entre plusieurs variables à partir d'observations. Dans cette section, nous abordons de manière non exhaustive les approches d'inférence causale rencontrées dans la littérature. Nous commençons par formaliser le problème d'inférence causale en introduisant les notions de système et de schéma causal ainsi que les outils graphiques de représentation de schéma causal. Nous présentons ensuite les approches d'inférence causale.

2.2.1 Formalismes du problème de causalité

Considérons un système $\mathbf{\Lambda} = (\Lambda_1, \dots, \Lambda_D)$ de D composantes. Chez l'individu $i = 1, \dots, n$, notons $\mathbf{\Lambda}_i = (\Lambda_{i1}, \dots, \Lambda_{iD})$ le système et \mathbf{Y}_{id} le marqueur de la composante d , possiblement mesuré de manière répétée chez l'individu i : $\mathbf{Y}_{id} = (Y_{i1}, \dots, Y_{idn_{id}})$ où n_{id} est le nombre de mesures répétées du marqueur \mathbf{Y}_{id} chez l'individu i . Posons $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iD})$ l'observation du système $\mathbf{\Lambda}$ chez l'individu i .

Supposons que le système $\mathbf{\Lambda}$ est régi par un mécanisme qui induit une structure de relations causales entre ses D composantes.

Le problème d'inférence causale consiste à inférer le mécanisme causal entre toutes les composantes ou un sous ensemble de composantes du système $\mathbf{\Lambda}$ à partir d'observations des marqueurs \mathbf{Y}_{id} ($d = 1, \dots, D$) dans une population de n individus.

2.2.1.1 Schéma causal

Lorsqu'on s'intéresse au mécanisme du système $\mathbf{\Lambda}$ sous-jacent aux données observées, il est important de délimiter le système et de représenter les hypothèses causales du système. Concrètement il s'agit de définir un schéma causal qui identifie les composantes du système $\mathbf{\Lambda}$ et l'ensemble de relations asymétriques entre les composantes du système.

Ces relations asymétriques découlent d'hypothèses faibles et fortes [Wermuth et Lauritzen, 1990]. Une hypothèse forte relative aux composantes Λ_d et $\Lambda_{d'}$ traduit l'absence d'une influence directe entre les composantes Λ_d et $\Lambda_{d'}$ alors qu'une hypothèse faible sur le lien Λ_d vers $\Lambda_{d'}$ traduit l'existence possible d'une influence directe de la composante Λ_d sur la composante $\Lambda_{d'}$ mais dont la force de la relation est à quantifier à partir des données [Bollen et Pearl, 2013].

Dans la modélisation causale, l'intérêt est souvent porté sur deux composantes Λ_1 (variable réponse) et Λ_2 (variable d'exposition). Il est important de situer ces deux com-

posantes dans un système Λ [Commenges et Gégout-Petit, 2009]. La notion de système est nécessaire pour d'une part identifier les composantes suffisantes dans l'étude de la relation causale entre Λ_1 et Λ_2 et d'autre part pour décomposer l'effet causal de la composante Λ_2 sur la composante Λ_1 en effet direct et effet indirect via d'autres composantes.

2.2.1.2 Représentation graphique

Une fois le système Λ et les relations entre les composantes définis dans un schéma causal, les graphes sont utilisés pour représenter le schéma causal [Lauritzen et Wermuth, 1989; Didelez, 2008].

Graphe acyclique direct (DAG)

Le DAG (Directed Acyclic Graph) est un graphe orienté utilisé pour représenter un schéma causal. Un DAG est composé de nœuds $V = \{\Lambda_d, d = 1, \dots, D\}$ reliés entre eux par des arcs orientés (Figure 2.1).

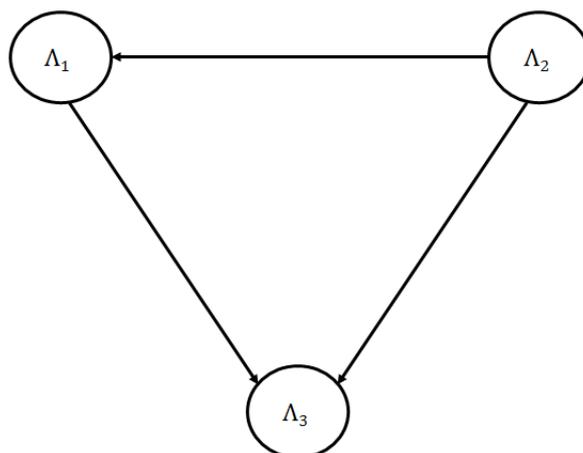


FIGURE 2.1 – DAG représentant un système Λ de trois composantes Λ_1 , Λ_2 et Λ_3 observées une seule fois.

Chaque nœud représente une composante du système Λ et un arc orienté du nœud Λ_d vers le nœud $\Lambda_{d'}$ indique un effet causal direct de la composante Λ_d sur la composante $\Lambda_{d'}$. Les nœuds Λ_d et $\Lambda_{d'}$ sont appelés respectivement « nœud ascendant » et « nœud descendant ».

L'ensemble $\mathcal{A}_d = \{\Lambda_{d^*}, d^* \in \{1, \dots, D\}, d^* \neq d\}$ tel qu'il existe un arc de Λ_{d^*} vers Λ_d est appelé « ascendants » ou « parents » du nœud Λ_d . De manière analogue, l'ensemble $\mathcal{D}_d = \{\Lambda_{d^*}, d^* \in \{1, \dots, D\}, d^* \neq d\}$ tel qu'il existe un arc de Λ_d vers Λ_{d^*} est appelé

« descendants » ou « fils » du nœud Λ_d [Pearl, 2009]. Un « chemin » entre deux nœuds Λ_d et Λ_{d^*} est une succession d'arcs entre ces deux nœuds.

Par exemple dans la figure 2.1, les nœuds Λ_1 et Λ_2 sont les ascendants du nœud Λ_3 et le nœud Λ_1 est un descendant de Λ_2 . $\Lambda_2 \rightarrow \Lambda_1 \rightarrow \Lambda_3$ est un chemin du nœud Λ_2 vers le nœud Λ_3 .

Les composantes représentées dans un DAG sont des variables aléatoires observées à des temps souvent différents. Cependant, le temps n'est pas explicitement pris en compte et modélisé dans les modèles statistiques qui découlent du DAG. Arjas et Parner [2004] ont souligné l'importance d'expliciter le temps dans un modèle de causalité, pas uniquement parce que la cause doit précéder l'effet mais aussi parce que souvent les durées entre les différents évènements font partie intégrante du problème causal.

Il est toutefois possible de représenter dans un DAG des relations asymétriques entre les composantes d'un système Λ répétées aux temps discrets t . Par exemple la Figure 2.2 représente un système dynamique Λ de trois composantes Λ_1 , Λ_2 et Λ_3 , telles que la composante Λ_3 au temps t peut influencer directement les composantes Λ_1 et Λ_2 aux temps suivant $t + 1$. Cette dernière représentation permet de préciser l'idée d'ordre entre la cause et l'effet.

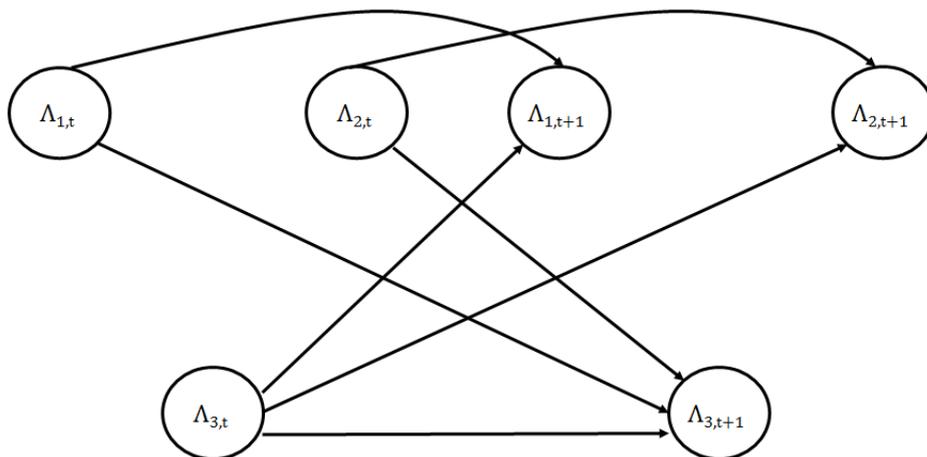


FIGURE 2.2 – DAG représentant un système Λ de trois composantes Λ_1 , Λ_2 , Λ_3 observées de manière répétée dans le temps. Les instances du système au temps t peuvent influencer les instances du système au temps $t + 1$.

Le graphe de dépendance locale (LDG)

Contrairement au DAG, le graphe de dépendance locale (LDG acronyme de Local Dependency Graph) propose une représentation d'un schéma causal qui tient compte de l'aspect dynamique et stochastique des composantes du système Λ [Didelez, 2008]. Dans un LDG (Figure 2.3), un nœud indique un processus stochastique et un arc d'un nœud Λ_d vers un nœud $\Lambda_{d'}$ indique l'influence causale du processus Λ_d sur le processus $\Lambda_{d'}$ quantifiée par l'effet du processus Λ_d sur le changement instantané du processus $\Lambda_{d'}$. Le formalisme du LDG, tient compte à la fois de l'aspect dynamique de la causalité et de l'aspect philosophique selon lequel l'effet causal direct d'une composante Λ_d sur une composante $\Lambda_{d'}$ doit être mesuré par le changement que la composante Λ_d induit sur la dynamique de la composante $\Lambda_{d'}$.

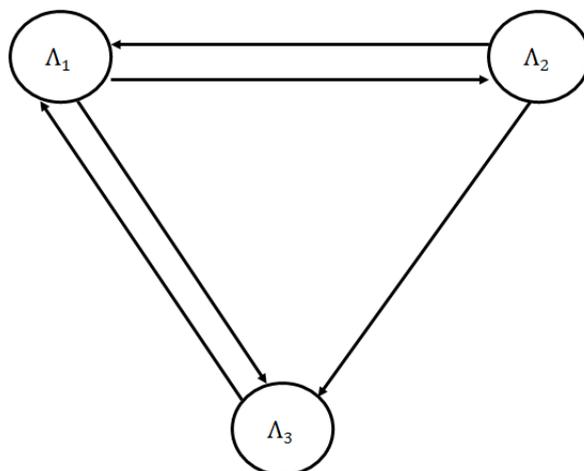


FIGURE 2.3 – Graphe de dépendance locale représentant un système Λ de trois composantes Λ_1 , Λ_2 , Λ_3 ; dans le LDG, les composantes du système Λ sont stochastiques et un arc orienté d'une composante $\Lambda_{d'}$ vers une composante Λ_d indique l'effet de la composante $\Lambda_{d'}$ sur le changement instantané de la composante Λ_d .

Dans l'exemple de la Figure 2.3 la composante Λ_1 cause un changement instantané des composantes Λ_2 et Λ_3 qui à leur tour cause aussi un changement de la composante Λ_1 . De plus la composante Λ_2 cause le changement de la composante Λ_3 .

Démarche de modélisation causale

Les graphes offrent des représentations graphiques d'un schéma causal. Ces représentations graphiques s'insèrent dans une démarche de modélisation causale pour laquelle nous avons identifié trois étapes :

- définition du schéma causal,
- choix de l'approche d'inférence causale,
- traduction du schéma causal en modèle. statistique

Le schéma causal représenté dans un graphe est traduit en un modèle statistique suivant l'approche d'inférence causale choisie. Dans notre revue de la littérature, nous avons identifié trois approches de causalité. Il s'agit de l'approche contrefactuelle [Rubin, 1974], l'approche de Granger [Granger, 1969] ou approche prédictive de causalité et enfin l'approche mécaniste ou stochastique de la causalité [Mykland, 1986].

Le graphe de dépendance locale est généralement traduit en système d'équations différentielles dans l'approche mécaniste de causalité. Le DAG est quant à lui traduit en modèle de régression multiple, en réseau Bayésien, en modèle d'équations structurelles dans l'approche de causalité de Granger [Granger, 1969] ou en modèle marginal structurel dans l'approche contrefactuelle.

Dans les sections suivantes nous abordons chaque approche de causalité, les modèles statistiques qui s'y rattachent, puis les avantages et les limites de l'approche.

2.2.2 Approche contrefactuelle de causalité

2.2.2.1 Présentation de l'approche contrefactuelle de causalité

Considérons deux composantes Λ_1 et Λ_2 d'un système Λ mesurées respectivement par les variables Y_i et X_i chez l'individu i dans une population I .

L'approche contrefactuelle ou modèle à réponse potentielle [Holland et al., 1985; Rubin, 1974] se base sur le concept de réponse potentielle Y_i^x de l'individu i pour quantifier l'effet causal de la variable X_i sur la variable Y_i au sein de l'unité expérimentale i (l'individu) dans une population I . Y_i^x représente la variable réponse de l'individu i si la variable X_i avait eu pour valeur x .

Dans un souci de clarté, nous allons considérer uniquement le cas où la variable X_i est binaire. Mais la méthodologie est la même pour une variable X_i catégorielle ou continue. La variable X_i peut désigner un traitement, avec $X_i = 1$ si l'individu i est traité et $X_i = 0$ sinon. La variable Y_i peut désigner un marqueur d'une maladie.

L'effet du traitement sur le marqueur Y_i est mesuré en comparant les deux variables potentielles Y_i^1 et Y_i^0 .

Plusieurs échelles sont utilisées pour cette comparaison, comme, par exemple l'échelle de la différence $c_i = Y_i^1 - Y_i^0$ ou l'échelle de rapport $c_i = Y_i^1/Y_i^0$.

2.2.2.2 Problème fondamentale de l'approche contrefactuelle de causalité

La comparaison des variables potentielles Y_i^1 et Y_i^0 est impossible dans la réalité. Car une seule des deux variables Y_i^1 et Y_i^0 est observée. C'est la variable factuelle, l'autre étant la variable contrefactuelle.

Cette situation est désignée par « problème fondamental de l'inférence causale » de l'approche contrefactuelle Holland et al. [1985].

A défaut, l'approche contrefactuelle de causalité utilise toute l'information dans la population I pour construire un estimateur non biaisé et consistant de l'effet moyen $C = \mathbb{E}_I(c_i)$ du traitement ($X = 1$) relativement au placebo ($X = 0$). Par exemple, dans l'échelle de la différence, L'effet moyen C s'écrit :

$$C = \mathbb{E}_I[Y_i^1] - \mathbb{E}_I[Y_i^0] \quad (2.11)$$

De par nature, l'approche contrefactuelle infère l'effet causal moyen dans une population.

2.2.2.3 Hypothèses et biais dans l'approche contrefactuelle de causalité

L'approche contrefactuelle de causalité repose sur les hypothèses suivantes :

- Il est possible d'associer à chaque individu i des variables réponses potentielles Y_i^0 et Y_i^1 conditionnées sur la variable X ,
- La variable d'intérêt Y_i coïncide avec les variables réponses potentielles Y_i^0 et Y_i^1 : $Y \equiv Y_i^0$ et $Y \equiv Y_i^1$,
- il n'existe pas de variable Z_i non observée qui confonde l'effet de X_i sur Y_i ,
- une indépendance (« strongly ignorability of treatment assignment assumption ») entre la variable X_i et les variables réponses potentielles : $X_i \perp\!\!\!\perp (Y_i^0, Y_i^1)$.

L'hypothèse d'indépendance entre la variable X_i et chaque variable potentielle (« strongly ignorability of treatment assignment assumption ») est centrale dans l'approche contrefactuelle de causalité [Rosenbaum et Rubin, 1983].

À cause du problème fondamental de l'inférence causale, il est uniquement possible d'obtenir l'estimation non biaisée de la quantité

$$A = \mathbb{E}_I [Y_i^1 | X_i = 1] - \mathbb{E}_I [Y_i^0 | X_i = 0].$$

Sous l'hypothèse d'indépendance entre X_i et les variables réponses potentielles, $X_i \perp\!\!\!\perp (Y_0, Y_1)$,

on obtient que

$$\begin{aligned}\mathbb{E}_I [Y_i^0 | X_i = 0] &= \mathbb{E}_I [Y_i^0] \\ \mathbb{E}_I [Y_i^1 | X_i = 1] &= \mathbb{E}_I [Y_i^1]\end{aligned}$$

avec $\mathbb{E}_I(\cdot)$ la fonction espérance.

Un estimateur non biaisé de A est donc aussi un estimateur consistant de l'effet causal moyen C [Holland et al., 1985].

Le quantité A est un paramètre d'association et mesure l'association entre les variables X et Y alors que le paramètre C mesure l'effet causal de la variable X sur la variable Y au sens contrefactuel. Le paramètre A est toujours estimable à partir des observations des individus I .

Sous l'hypothèse que X (cause) est indépendante des variables potentielles réponses Y^X , les paramètres A et C sont égaux et l'effet causal C de la variable X sur la variable Y se déduit de la mesure d'association A entre X et Y .

L'hypothèse d'indépendance est satisfaite dans les essais randomisés mais ne l'est pas souvent dans les études en population à cause des variables de confusion non contrôlées par la sélection. Ainsi dans les études en population, l'estimateur de l'effet causal par l'approche contrefactuelle de causalité moyen est souvent biaisé [Holland et al., 1985].

2.2.2.4 Extension de l'approche contrefactuelle de causalité aux études en population

L'approche contrefactuelle s'est heurtée au problème fondamental de l'inférence causale qui limite son utilisation exclusivement aux essais randomisés et pour inférer uniquement un effet causal moyen dans la population.

Des travaux ont tenté d'étendre l'utilisation de l'approche contrefactuelle aux études en population [Rosenbaum et Rubin, 1983; Robins et al., 2000; Hernán et al., 2002]. L'idée générale est de remplacer l'hypothèse forte d'indépendance entre X_i et les variables réponses potentielles (Y_i^0, Y_i^1) par une hypothèse moins forte d'indépendance conditionnelle.

Elle suppose qu'il existe certaines variables de confusion \mathbf{Z}_i observées, telles que conditionnellement à \mathbf{Z}_i , la variable X_i est indépendante des variables potentielles (Y_i^0, Y_i^1) .

En présence de variable de confusion \mathbf{Z}_i , les paramètres d'association et d'effet causal

deviennent respectivement

$$A = \mathbb{E} \left[\mathbb{E}(Y_i^1 | X_i = 1, \mathbf{Z}_i = \mathbf{z}_i) - \mathbb{E}(Y_i^0 | X_i = 0, \mathbf{Z}_i = \mathbf{z}_i) \right] \quad (2.12)$$

et

$$C = \mathbb{E} \left[\mathbb{E}(Y_i^1 | \mathbf{Z}_i = \mathbf{z}_i) - \mathbb{E}(Y_i^0 | \mathbf{Z}_i = \mathbf{z}_i) \right]. \quad (2.13)$$

Sous l'hypothèse d'indépendance conditionnelle, les paramètres d'association et d'effet causal sont égaux.

Rosenbaum et Rubin [1983] ont montré l'existence de fonctions des variables de confusion \mathbf{Z}_i à valeurs réelles $p(\mathbf{Z}_i)$ telles que l'indépendance conditionnelle entre X_i et les variables potentielles sachant les variables \mathbf{Z}_i entraîne une indépendance conditionnelle entre X et les variables potentielles (Y_i^0, Y_i^1) sachant la fonction $p(\mathbf{Z}_i)$ ou sachant un estimateur $\hat{p}(\mathbf{Z}_i)$:

$$X_i \perp\!\!\!\perp (Y_i^0, Y_i^1) | \mathbf{Z}_i \equiv X_i \perp\!\!\!\perp (Y_i^0, Y_i^1) | \hat{p}(\mathbf{Z}_i) \quad (2.14)$$

Ces fonctions sont appelées « balancing score » [Rosenbaum et Rubin, 1983]. Dans le cas d'une variable binaire X_i , Rosenbaum et Rubin [1983] ont proposé un estimateur non-biaisé et consistant de l'effet causal de X_i sur Y_i en utilisant comme « balancing score » le score de propension $p_i = \mathbb{P}[X_i = 1 | \mathbf{Z}_i]$, c'est-à-dire la probabilité que l'individu i a d'être traité sachant ses variables de confusion \mathbf{Z}_i .

L'extension du modèle contrefactuel apportée par Rosenbaum et Rubin [1983] est adaptée pour une variable (cause) X_i binaire et non-dépendante du temps. Robins et al. [2000] ont proposé une nouvelle classe d'estimateur qui généralise celle de Rosenbaum et Rubin [1983] au cas d'une variable (cause) X_i continue et dépendant du temps.

Robins et al. [2000] ont remplacé la fonction de score de propension de Rosenbaum et Rubin [1983] par la fonction $w_i = 1/\mathbb{P}[X_i = x | \mathbf{Z}_i]$ qui est l'inverse de la probabilité d'observer chez l'individu i la modalité $X_i = x$ sachant les variables de confusion \mathbf{Z}_i . Cette classe d'estimateur est connue dans la littérature sous l'acronyme IPTW pour « Inverse Probability-of-Treatment Weighted ». Robins et al. [2000] ont formalisé leur extension dans le modèle structurel marginal.

2.2.2.5 Modèle marginal structurel

Dans l'approche de causalité contrefactuelle, Robins et al. [2000] ont proposé le modèle marginal structurel MSM (« Marginal Structural Model ») pour estimer sans biais l'effet

de traitement dépendant du temps X_{it} sur un marqueur d'intérêt Y_i mesuré une fois, et en présence de facteur de confusion Z_{it} dépendant du temps.

La méthodologie de Robins consiste à proposer deux modèles marginaux paramétriques pour la variable contrefactuelle $Y_i^{\bar{x}}$ et la variable observée Y_i (équations (2.15)) puis de déduire l'estimation non biaisée de l'effet moyen du traitement (ϕ_1) à partir de l'estimation du paramètre d'association moyenne ϕ'_1 .

$$\begin{cases} \text{logit}(\mathbb{P}[Y^{\bar{x}} = 1]) & = \phi_0 + \phi_1 \text{cum}(\bar{x}_t) \\ \text{logit}(\mathbb{P}[Y = 1 | \bar{X} = \bar{x}]) & = \phi'_0 + \phi'_1 \text{cum}(\bar{x}_t) \end{cases} \quad (2.15)$$

avec $\text{cum}(\bar{x}_t)$ l'exposition cumulée au traitement X_t associée au profil d'exposition $\bar{X}_t = \bar{x}_t$ (c'est-à-dire l'historique de la variable X_t jusqu'au temps t), ϕ_0 et ϕ'_0 les intercepts respectifs dans les modèles de causalité et d'associations.

L'association moyenne entre le traitement $X_{i,t}$ et la variable Y_i étant confondue par la variable $\mathbf{Z}_{i,t}$, le paramètre d'association ϕ'_1 est un estimateur biaisé de ϕ_1 . Robins et al. [2000] ont proposé de corriger ce biais par la méthode de pondération IPTW (« Inverse probability of treatment weighted »). A chaque individu i est attribué un poids sw_i qui est l'inverse de la probabilité qu'il a de recevoir son traitement sachant la variable de confusion $\mathbf{Z}_{i,t}$.

$$sw_i = \frac{\prod_{t=0}^T \mathbb{P}[X_{i,t} = x_{i,t} | \bar{X}_{i,t-1} = \bar{x}_{i,t-1}]}{\prod_{t=0}^T \mathbb{P}[X_{i,t} = x_{i,t} | \bar{X}_{i,t-1} = \bar{x}_{i,t-1}; \bar{\mathbf{Z}}_{i,t} = \bar{\mathbf{z}}_{i,t}]} \quad (2.16)$$

où T est le nombre de temps d'observation.

Des sous-modèles marginaux paramétriques sont utilisés à chaque temps d'observation t pour modéliser les probabilités nécessaires à l'estimation des poids sw_i . Par exemple la régression logistique pour données répétées binaires est utilisée lorsque le traitement est binaire. Dans ce cas les poids sw_i sont estimés par :

$$s\hat{w}_i = \frac{\prod_{t=0}^T \hat{p}_{i,t}^* (1 - \hat{p}_{i,t}^*)^{(1-x_{i,t})}}{\prod_{t=0}^T \hat{p}_{i,t} (1 - \hat{p}_{i,t})^{(1-x_{i,t})}} \quad (2.17)$$

où $\hat{p}_{i,t}^* = \hat{\mathbb{P}}(X_{i,t} = 1 | \bar{X}_{i,t-1} = \bar{x}_{i,t-1})$ et $\hat{p}_{i,t} = \hat{\mathbb{P}}(X_{i,t} = 1 | \bar{X}_{i,t-1} = \bar{x}_{i,t-1}; \bar{\mathbf{Z}}_{i,t} = \bar{\mathbf{z}}_{i,t})$.

Hernán et al. [2002] ont étendu le modèle marginal structurel pour estimer l'effet causal d'un traitement dépendant du temps $X_{i,t}$ sur une variable répétée $Y_{i,t}$ en présence de variable de confusion $\mathbf{Z}_{i,t}$ dépendant du temps.

Estimation du modèle marginal structurel

L'estimation du modèle MSM est faite en deux étapes. Premièrement, les paramètres des modèles relatifs au calcul des poids sw_i sont estimés par maximum de vraisemblance ou par la méthode des moindres carrés. Ensuite les poids estimés $s\hat{w}_i$ sont utilisés comme poids dans un modèle linéaire généralisé potentiellement ajusté sur des variables explicatives et estimé par la méthode des équations d'estimation généralisées GEE (« Generalized Estimating Equations »).

Applications du modèle marginal structurel

Hernán et al. [2002] ont utilisé leur approche pour estimer l'effet causal du traitement au zidovudine sur les cellules CD4 dans le VIH. Dans la maladie d'Azheimer, Dutcher et al. [2014] ont utilisé le modèle MSM pour modéliser l'effet d'un traitement préventif de la démence sur le déclin fonctionnel mesuré par l'échelle ADL (« Activities of Daily Living ») possiblement confondu par le niveau cognitif mesuré au cours du temps. Dans cette application, Dutcher et al. [2014] ont montré qu'après ajustement sur le niveau cognitif courant, le traitement préventif de la démence n'était pas associé au déclin fonctionnel.

Limites du modèle marginal structurel

Le modèle marginal présente certaines limites. D'abord, il n'est pas adapté dans le cas où, à un temps d'observation donné, tous les individus reçoivent le même traitement sachant leur variable de confusion Z_t . Dans ce cas les poids sw_i ne sont pas estimables [Robins et al., 2000]. Il n'est pas adapté non plus dans le cas de traitement observé avec des erreurs de mesure, ce qui est fréquent quand X n'est plus un traitement mais un autre marqueur collecté dans une cohorte.

Une autre limite du modèle marginal structurel est liée à la procédure d'estimation en deux étapes. Une mauvaise spécification des sous-modèles pour le calcul des probabilités dans l'expression des poids sw_i peut entraîner un biais sur l'estimation de l'effet causal [Robins et al., 2000].

2.2.3 Causalité de Granger

2.2.3.1 Présentation de la causalité de Granger

L'approche de causalité dite de Granger et encore appelée approche prédictive de la causalité, est formellement définie par Granger [1969] mais l'idée remonte aux travaux de Wiener [1956]. L'idée générale de cette approche est que l'effet causal d'une série

stochastique (\mathbf{X}_t) sur une série stochastique (\mathbf{Y}_t) est mesurée par la capacité prédictive de la série (\mathbf{X}_t) sur la série (\mathbf{Y}_t) . Pour deux séries stochastiques, (\mathbf{X}_t) et (\mathbf{Y}_t) , de filtrations respectives $\mathcal{F}_t^x = \{X_\tau, \tau = 0, \dots, t-1\}$ et $\mathcal{F}_t^y = \{Y_\tau, \tau = 0, \dots, t-1\}$, la série (\mathbf{X}_t) cause au sens de Granger la série (\mathbf{Y}_t) si l'erreur quadratique moyenne de prédiction de (\mathbf{Y}_t) conditionnée aux valeurs passées de (\mathbf{X}_t) et de (\mathbf{Y}_t) est inférieure à l'erreur quadratique moyenne de prédiction de (\mathbf{Y}_t) conditionnée uniquement sur le passé de (\mathbf{Y}_t) , soit :

$$\mathbb{E} \left[\left(Y_t - \hat{Y}_t | \mathcal{F}_t^y, \mathcal{F}_t^x \right)^2 \right] < \mathbb{E} \left[\left(Y_t - \hat{Y}_t | \mathcal{F}_t^y \right)^2 \right] \quad (2.18)$$

La formulation de la causalité au sens de Granger explicite d'une part la condition que, dans une relation causale, la cause doit précéder l'effet et la loi de probabilité de l'effet est affectée par les valeurs passées de la cause (concept du philosophe David Hume). D'autre part elle nécessite le calcul des prédictions $\hat{Y}_t | \mathcal{F}_t^y$ et $\hat{Y}_t | \mathcal{F}_t^y, \mathcal{F}_t^x$ respectivement à partir des distributions de probabilités conditionnelles $f_{Y_t | \mathcal{F}_t^y}$ et $f_{Y_t | \mathcal{F}_t^y, \mathcal{F}_t^x}$ de la série (\mathbf{Y}_t) . La relation (2.18) signifie que les prédictions $\hat{Y}_t | \mathcal{F}_t^y, \mathcal{F}_t^x$ de Y_t sont meilleures que les prédictions $\hat{Y}_t | \mathcal{F}_t^y$. Bien que formulée dans le cadre de séries temporelles, l'approche de causalité de Granger peut être facilement utilisée dans le cas de variables X et Y non indexées par le temps avec la même idée que les prédictions de Y à partir de la distribution conditionnelle $f_{Y|X}$ sont meilleures que celles obtenues à partir de la distribution marginale f_Y .

L'approche de causalité de Granger peut être retrouvée dans le modèle de régression classique, le modèle à équations structurelles et les réseaux Bayésiens s'ils sont utilisés pour inférer une relation causale.

2.2.3.2 Régression linéaire multiple

Une situation très fréquente est de s'intéresser dans un système $\mathbf{\Lambda}$ de D composantes à l'effet d'une composante Λ_2 sur une composante Λ_1 à partir des observations Y_1, \dots, Y_D du système. Les variables Y_1, \dots, Y_D peuvent être scindées en trois parties : $\{Y_1\}$, $\{Y_2\}$ et $\{Z_1, \dots, Z_{D-2}\} = \{Y_3, \dots, Y_D\}$.

La régression linéaire multiple suivante est traditionnellement utilisée pour estimer l'effet de la variable Y_2 sur la variable dépendante Y_1 .

$$Y_1 = \beta_1 + \beta_2 Y_2 + \beta_3 Z + \epsilon_1, \quad (2.19)$$

avec ϵ_1 l'erreur de mesure de la variable Y_1 , β_1 l'intercept, et β_2, β_3 les effets fixes des variables indépendantes Y_2 et $Z \in \{Y_3, \dots, Y_D\}$.

Le modèle de régression multiple (2.19) repose sur les hypothèses suivantes :

ÉTAT DE L'ART

- la variable dépendante Y_1 n'influence pas directement les variables indépendantes Y_2, Z_1, \dots, Z_{D-2} (hypothèse forte),
- les variables Y_2, Z influencent possiblement la variable dépendante Y_1 (hypothèse faible),
- les variables Y_2, Z sont mesurées sans erreurs .

Ces hypothèses entraînent un schéma causal représenté par le DAG de la Figure 2.4.

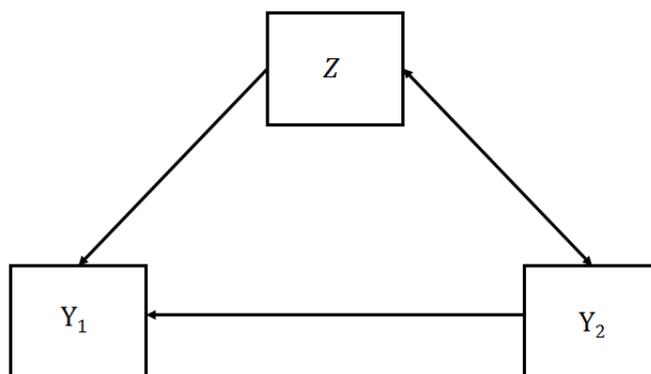


FIGURE 2.4 – DAG correspondant au schéma causal modélisé dans un modèle de régression linéaire multiple où Y_1 est la variable dépendante, Z une variable de confusion et Y_2 la variable indépendante.

Le paramètre β_2 a une interprétation causale si les variables d'ajustement sont judicieusement choisies parmi \mathbf{Z} de sorte que l'effet causal de y_2 sur Y_1 ne soit pas confondu par ces variables.

Dans le cas où la variable dépendante Y_1 est répétée chez l'individu i , $\mathbf{Y}_{i1} = (Y_{ij1})_{j=1, n_i}$ et que les variables indépendantes Y_2 et Z sont fixes, le modèle linéaire mixte décrit dans la section 2.1 peut être utilisé de manière analogue à la régression linéaire multiple pour inférer l'effet causal de la variable Y_2 sur la variable Y_1 tout en tenant compte des corrélations entre les mesures répétées de la variable Y_1 chez l'individu i . Le modèle linéaire mixte multivarié étend le modèle linéaire mixte univarié au cas de plusieurs variables dépendantes \mathbf{Y}_{1i} mais il n'est pas adapté pour inférer des relations causales entre ces marqueurs dépendants du temps (voir section 2.1.2.6).

2.2.3.3 Réseau Bayésien

Pearl [2003] a proposé le modèle de réseau Bayésien comme alternative à la démarche

traditionnelle de traduction d'un DAG en modèle de régression linéaire multiple. Un réseau Bayésien est une paire $B = (G, \Theta)$ où G est un DAG et Θ l'ensemble des paramètres du réseau permettant de définir une unique distribution jointe de probabilité des variables Y_1, \dots, Y_D qui sont les observations respectives des composantes $\Lambda_1, \dots, \Lambda_D$ du système Λ [Friedman et al., 1997]. Un réseau Bayésien modélise quantitativement à l'aide de probabilités conditionnelles les effets causaux entre les variables aléatoires. Les hypothèses d'indépendance conditionnelle encapsulées dans le DAG G permettent une simplification de la distribution jointe de probabilité :

$$f_Y(Y_1 = y_1, \dots, Y_D = y_D; \theta) = \prod_d f(Y_d = y_d | \mathcal{A}_d; \theta) \quad (2.20)$$

avec \mathcal{A}_d l'ensemble des nœuds ascendants du nœud d et θ les paramètres du modèle de réseau Bayésien.

À partir de la distribution jointe de probabilité du réseau de variables aléatoires, la distribution marginale de deux variables quelconques (Y_1 et Y_2) peut être calculée en intégrant la densité jointe de probabilité sur les autres variables restant :

$$f_{Y_1, Y_2}(y_1, y_2; \theta) = \int_{\mathbb{R}^{(D-2)}} f_Y(y_1, y_2, y_3, \dots, y_D; \theta) dy_3 \dots dy_D \quad (2.21)$$

La distribution de la variable aléatoire Y_1 conditionnelle à la variable aléatoire Y_2 peut alors s'écrire :

$$f_{Y_1|Y_2}(y_1, y_2; \theta) = \frac{\int_{\mathbb{R}^{(D-2)}} f_Y(y_1, y_2, y_3, \dots, y_D; \theta) dy_3 \dots dy_D}{\int_{\mathbb{R}^{(D-1)}} f_Y(y_1, y_2, y_3, \dots, y_D; \theta) dy_1 dy_3 \dots dy_D} \quad (2.22)$$

La comparaison des densités de probabilités $f_{Y_1|Y_2}(y_1, y_2; \theta)$ et $f_{Y_1}(y_1; \theta)$ par exemple dans l'échelle de la différence $f_{Y_1|Y_2}(y_1; \theta) - f_{Y_1}(y_1; \theta)$ permet de quantifier l'effet global de la variable aléatoire Y_2 sur la variable aléatoire Y_1 .

En reprenant le cas du DAG de la Figure 2.1 de la section 2.2.1.2, les conditions d'indépendance représentées dans le DAG entraînent une simplification de la distribution de probabilité jointe des marqueurs Y_1 , Y_2 et Y_3 mesurant les composantes Λ_1 , Λ_2 et Λ_3 , respectivement :

$$f_Y(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3; \theta) = f_{Y_3|Y_2, Y_1}(y_3; \theta) \times f_{Y_2|Y_1}(y_2; \theta) \times f_{Y_1}(y_1; \theta)$$

La densité conditionnelle de la variable aléatoire Y_1 à la variable Y_2 s'écrit alors :

$$f_{Y_1|Y_2}(y_1, y_2; \theta) = \frac{\int_{\mathbb{R}} f_{Y_3|Y_2, Y_1}(y_3; \theta) \times f_{Y_2|Y_1}(y_2; \theta) \times f_{Y_1}(y_1; \theta) dy_3}{\int_{\mathbb{R}^2} f_{Y_3|Y_2, Y_1}(y_3; \theta) \times f_{Y_2|Y_1}(y_2; \theta) \times f_{Y_1}(y_1; \theta) dy_1 dy_3} \quad (2.23)$$

Le même schéma causal peut être utilisé pour calculer les autres densités de probabilité conditionnelles ; par exemple la densité de probabilité conditionnelle de la variable Y_2 par rapport à la variable aléatoire Y_1 pour investiguer l'effet causal de la variable Y_1 sur la variable Y_2 .

Une différence majeure par rapport à la régression multiple est qu'ici, toutes les variables du réseau peuvent être mesurées avec des erreurs de mesures.

Réseau Bayésien dynamique

Dans le réseau Bayésien, les composantes sont des variables observées, chacune à un seul temps d'observation. L'analyse causale ne tient donc pas compte explicitement du temps. Or l'importance du temps et la conséquence de sa non prise en compte dans la démarche d'analyse causale sont bien connues de la littérature [Aalen et al., 2016; Arjas et Parner, 2004].

Le réseau Bayésien dynamique (DBN pour « Dynamic Bayesian Network ») est l'extension naturelle du réseau Bayésien qui intègre le temps dans la modélisation causale en indexant par le temps une succession de réseaux Bayésiens [Kjærulff, 1995]. Le DBN est une séquence de réseaux Bayésiens définis en temps discret t avec des interconnexions (arcs) entre des composantes de deux séquences de réseaux Bayésiens. Un arc inter-réseaux reliant deux nœuds $\Lambda_{d', (t-k)}$ et $\Lambda_{d, t}$ représente l'effet de la composante $\Lambda_{d'}$ prise au temps $t - k$ sur la composante Λ_{d} au temps t [Ghahramani et Hinton, 1996]. La Figure 2.2 de la section 2.2.1.2 est un exemple de DBN entre deux séquences consécutives.

Les modèles auto-régressifs d'ordre k , les modèles de Markov cachés (HMM), les réseaux de neurones, les « state-space models (SSM) » et les filtres de Kalman en particulier sont des réseaux Bayésiens dynamiques [Murphy et Russell, 2002; Ghahramani, 1998].

Le modèle auto-régressif d'ordre 1 par exemple est une formulation paramétrique du réseau Bayésien dynamique dans laquelle les hypothèses causales du DAG sont encapsulées dans une matrice de transition $\mathbf{A} = (a_{d'd})_{d, d' \in \{1, \dots, D\}^2}$:

$$\mathbf{Y}_{t+1} = \mathbf{A}\mathbf{Y}_t + \boldsymbol{\epsilon}_{t+1} \quad (2.24)$$

où $\boldsymbol{\epsilon}_{t+1}$ est le vecteur des erreurs de mesure indépendantes de matrice de variance $\boldsymbol{\Sigma}$: $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Dans la matrice \mathbf{A} , un coefficient $a_{d'd}$ non nul exprime l'influence de la composante d sur la dimension d' .

Dans le réseau Bayésien dynamique, la structure d'influence entre les composantes du réseau ne dépend pas du temps. Le TV-DBN (« Time-varying Dynamic Bayesian Network ») est une extension qui permet de modéliser des liens causaux dépendants du temps entre D variables Y_d ($d = 1, \dots, D$) observées de manière répétée aux temps discrets $t = 0, 1, \dots, T$ [Song et al., 2009]. Par exemple $\mathbf{Y}_{t+1} = \mathbf{A}(t)\mathbf{Y}_t + \boldsymbol{\epsilon}_t$ est la version TV-DBN du modèle auto-régressif d'ordre 1.

Applications du réseau Bayésien

Les réseaux Bayésiens ont été popularisés en apprentissage automatique (machine learning) et en intelligence artificielle [Korb et Nicholson, 2010]. Mais récemment ils ont trouvé quelques applications en biostatistique et notamment dans l'étude du vieillissement. Par exemple, Wang et al. [2014] ont utilisé les réseaux Bayésiens pour comparer la structure de connexions entre 8 régions d'intérêt dans le jeune âge (entre 20 ans et 28 ans) et dans le grand âge (entre 60 ans et 90 ans). Un nœud du réseau Bayésien représente le volume de substance grise d'une région d'intérêt. Les variables aléatoires du réseau sont supposées Gaussiennes. Dans chaque groupe de sujets, plusieurs réseaux Bayésiens sont estimés par maximum de vraisemblance et le meilleur réseau Bayésien est sélectionné à l'aide du critère de BIC. Wang et al. [2014] ont montré dans cette application que la connectivité entre les régions du cerveau était réduite chez les personnes âgées relativement aux jeunes. Notamment les liens cortex temporal inférieur et cortex pré-frontal médial, hippocampe droit et cortex temporal inférieur, ainsi que hippocampe droit et cortex cingulaire postérieur qui disparaissent dans le grand âge.

Limites du réseau Bayésien

Les réseaux Bayésien (BN, DBN, TV-DBN) permettent de traduire un DAG complexe en un modèle statistique dans lequel toutes les composantes du DAG sont supposées stochastiques. Ils sont plus souples par rapport à la régression linéaire. Cependant, les réseaux Bayésien (BN, DBN, TV-DBN) ont quelques limites.

Les réseaux Bayésiens modélisent les marqueurs Y_1, \dots, Y_D des composantes possiblement latentes $\Lambda_1, \dots, \Lambda_D$ du système $\boldsymbol{\Lambda}$. Parfois, un marqueur Y_d peut ne pas mesurer correctement la composante Λ_d du système $\boldsymbol{\Lambda}$ de sorte que la structure de causalité entre

les marqueurs Y_1, \dots, Y_D peut ne pas refléter la vraie structure de causalité entre les composantes $\Lambda_1, \dots, \Lambda_D$ du système $\mathbf{\Lambda}$. Cette situation est fréquente en psychométrie où d'ailleurs plusieurs marqueurs sont collectés pour mesurer une même quantité sous-jacente. Dans ces cas, l'inférence causale dans le système $\mathbf{\Lambda}$ devrait plutôt se faire directement sur les variables sous-jacentes (non bruitées par les erreurs de mesure) à partir des marqueurs observés.

Dans le réseau Bayésien dynamique le temps est pris en compte. Toutefois, les nœuds représentent des variables mesurées aux temps discrets $t = 0, \dots, T$ avec un pas de discrétisation non explicité. Aalen et al. [2016] ont montré qu'une discrétisation (avec un pas grossier) pouvait conduire à une inférence erronée de la structure de connexions entre les composantes du système $\mathbf{\Lambda}$. Par exemple, la discrétisation peut créer une relation entre deux composantes qui est pourtant inexistante au niveau infinitésimal.

Enfin, dans le réseau Bayésien, l'effet causal est quantifié par l'influence d'une composante sur la valeur d'une autre composante (à travers les marqueurs). De ce fait, les réseaux Bayésiens ne sont pas compatibles avec l'idée de l'approche stochastique de la causalité selon laquelle, l'effet causal d'une composante Λ_d sur une composante $\Lambda_{d'}$ est quantifiée par l'effet qu'elle produit sur le changement de la composante $\Lambda_{d'}$.

2.2.3.4 Modèle d'incrémentes linéaires

Le modèle d'incrémentes linéaires a été proposé par Diggle et al. [2007] comme technique d'imputation multiple de données longitudinales incomplètes $\mathbf{Y}(t)$, puis utilisé par Aalen et Gunnes [2010] pour faire de l'inférence causale. Il s'agit d'un modèle multivarié des incréments de D marqueurs $\tilde{\mathbf{Y}} = (\tilde{Y}_1, \dots, \tilde{Y}_D)$ qui sont collectés chez n individus aux temps $t = 0, \dots, T$. Ce modèle s'écrit :

$$\Delta\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{Y}}(t-1)\boldsymbol{\beta}(t) + \tilde{\boldsymbol{\epsilon}}(t), \quad (2.25)$$

avec $\tilde{\mathbf{Y}}(t)$ une matrice $n \times D$, où la $i^{\text{ème}}$ ligne représente les observations de D marqueurs chez l'individu $i = 1, \dots, n$, $\Delta\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{Y}}(t) - \tilde{\mathbf{Y}}(t-1)$ une matrice $n \times D$ des incréments de D marqueurs de n individus au temps t , $\boldsymbol{\beta}(t)$ une matrice $D \times D$ de paramètres et $\tilde{\boldsymbol{\epsilon}}(t)$ une matrice $n \times D$ des erreurs de mesures.

Les données complètes $\tilde{\mathbf{Y}}(t)$ sont supposées admettre une décomposition de Doob-Meyer :

$$\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{\Lambda}}(t) + \tilde{\mathbf{M}}(t),$$

avec $\tilde{\mathbf{M}}(t)$ une martingale d'espérance nulle et $\tilde{\mathbf{\Lambda}}(t)$ un processus déterministe appelé compensateur et défini par :

$$\tilde{\mathbf{\Lambda}}(\mathbf{0}) = \tilde{\mathbf{y}}_0, \quad \tilde{\mathbf{\Lambda}}(t) = \sum_{l=1}^t \tilde{\mathbf{Y}}(l-1)\boldsymbol{\beta}(l), \quad \text{pour } t \geq 1,$$

avec $\tilde{\mathbf{y}}_0$ le niveau initial de m marqueurs mesurés sans erreurs chez n individus.

À chaque temps d'observation t des données peuvent être manquantes chez l'individu i . Les marqueurs incomplets $\mathbf{Y}(t)$ sont modélisés en les reliant aux marqueurs complets $\tilde{\mathbf{Y}}(t)$ à l'aide de matrices diagonales d'observation $\mathbf{R}_0(t)$ et $\mathbf{R}(t)$: $\mathbf{Y}(t) = \mathbf{R}_0(t)\tilde{\mathbf{Y}}(t)$ et $\Delta\mathbf{Y}(t) = \mathbf{R}(t)\Delta\tilde{\mathbf{Y}}(t)$. La matrice $\mathbf{R}_0(t)$ possède 1 sur la $i^{\text{ème}}$ ligne si tous les marqueurs sont observés au temps t chez l'individu i et 0 sinon. La matrice $\mathbf{R}(t)$ possède 1 sur la $i^{\text{ème}}$ ligne si tous les marqueurs sont observés aux temps t et $t-1$ chez l'individu i et 0 sinon. Le modèle complet (modèle pour les données complètes et le modèle d'observation) s'écrit :

$$\begin{cases} \tilde{\mathbf{Y}}(t) = \tilde{\mathbf{\Lambda}}(t) + \tilde{\mathbf{M}}(t) \\ \Delta\tilde{\mathbf{Y}}(t) = \tilde{\mathbf{Y}}(t-1)\boldsymbol{\beta}(t) + \tilde{\boldsymbol{\epsilon}}(t) \\ \mathbf{Y}(t) = \mathbf{R}_0(t)\tilde{\mathbf{Y}}(t) \\ \Delta\mathbf{Y}(t) = \mathbf{R}(t)\Delta\tilde{\mathbf{Y}}(t) \end{cases} \quad (2.26)$$

Estimation du modèle d'incrément linéaires

A chaque temps discret t , l'estimation des paramètres $\boldsymbol{\beta}(t)$ est obtenue par la méthode des moindres carrées :

$$\hat{\boldsymbol{\beta}}(t) = \left(U^\top U \right)^{-1} U^\top \Delta\mathbf{Y}(t), \quad (2.27)$$

avec $U = \mathbf{R}(t)\mathbf{Y}(t-1)$.

L'estimateur non-biaisé des paramètres $\boldsymbol{\beta}(t)$ repose sur l'hypothèse DTIC (« Discrete-Time Independence Censoring ») selon laquelle, conditionnellement à l'historique \mathcal{F}_{t-1} des données complètes jusqu'au temps $t-1$ (inclus), l'incrément des données complètes $\Delta\tilde{\mathbf{Y}}(t)$ est indépendant de l'historique \mathcal{R}_{t-1} des matrices d'observation $\mathbf{R}_0(t)$ et $\mathbf{R}(t)$ jusqu'au temps $t-1$ (inclus), soit :

$$\mathbb{E} \left[\Delta\tilde{\mathbf{Y}}(t) \Big|_{\mathcal{F}_{t-1}, \mathcal{R}_{t-1}} \right] = \mathbb{E} \left[\Delta\tilde{\mathbf{Y}}(t) \Big|_{\mathcal{F}_{t-1}} \right] \quad (2.28)$$

L'hypothèse DTIC garantit que $\Delta\mathbf{Y}(t) = \mathbf{R}(t)\mathbf{Y}(t-1)\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t)$.

Les estimations non-biaisées du compensateur $\hat{\tilde{\mathbf{\Lambda}}}(t)$ et des données complètes $\hat{\tilde{\mathbf{Y}}}(t)$ sont ensuite calculées à partir des estimateurs $\hat{\boldsymbol{\beta}}(t)$ pour $t \geq 0$, de $\boldsymbol{\beta}(t)$.

Causalité dans le modèle d'incrément linéaires

Le modèle d'incrément linéaires initialement proposé pour faire de l'imputation peut être utilisé comme modèle d'inférence causale entre les composantes d'un système Λ .

xxxxxxx Certains problèmes de causalité sont en effet perçus comme des problèmes d'imputation de données manquantes. Par exemple, l'évaluation de l'effet d'un traitement binaire T sur un marqueur Y_i mesuré chez des patients $i = 1, \dots, n$ peut se faire par comparaison des quantités $Y_i^{T=1}$ et $Y_i^{T=0}$ qui sont respectivement les valeurs aléatoires du marqueur Y_i qui seraient observées si le patient i avait été traité ($T = 1$) ou non ($T = 0$). Dans l'échelle de la différence, on peut calculer $Y_i^{T=1} - Y_i^{T=0}$ sauf que dans la réalité, une seule des valeurs $Y_i^{T=1}$ et $Y_i^{T=0}$ est observée; l'autre étant manquante. Dans ce cas de figure, le modèle d'incrément d'une part permet de reconstruire les données complètes $\tilde{Y}_i = (Y_i^{T=0}, Y_i^{T=1})$ de tous les patients et d'autre part d'inférer les effets $\beta_{Y_{1T}}$ et $\beta_{Y_{0T}}$ du traitement T respectivement sur $Y^{T=1}$ et $Y^{T=0}$. La comparaison des paramètres $\beta_{Y_{1T}}$ et $\beta_{Y_{0T}}$ toutes choses égales par ailleurs permet d'inférer l'effet causal moyen du traitement.

Les données reconstruites peuvent aussi être utilisées ultérieurement dans un autre modèle d'inférence causale.

Limites du modèle d'incrément linéaires

A nouveau cette approche permet d'appréhender des structures de graphes complexes. De plus, elle traduit bien l'approche stochastique de la causalité dans laquelle toutes les composantes du système peuvent être mesurées avec des erreurs et l'effet causal d'une composante sur une autre est quantifié par le changement qu'elle produit sur la composante d'intérêt.

Cependant, le modèle d'incrément linéaires présente quelques limites. Premièrement, à chaque temps d'observation, $D \times D$ paramètres d'interaction entre les composantes du système d'intérêt sont à estimer et un nombre conséquent de données complètes entre deux temps consécutifs est requis pour l'estimation des paramètres à chaque temps. Aussi, les paramètres $\beta(t)$ estimés aux temps d'observation ne sont pas reliés entre eux. Cette discontinuité des paramètres du modèle $\beta(t)$ est discutable. Deuxièmement, l'estimation des paramètres par la méthode des moindres carrés constitue aussi une limite car l'estimateur des moindres carrés est biaisé dans le cas de données très asymétriques. Troisièmement, dans le modèle d'incrément linéaires, l'hétérogénéité entre les sujets n'est pas modélisée. Cela peut induire des biais sur les paramètres marginaux $\beta(t)$ comme c'est le cas dans les études en population très hétérogène. Enfin, la causalité inférée dans le modèle d'incrément linéaires est à l'échelle des observations complètes et non à l'échelle de processus

non bruités.

2.2.3.5 Modèle à équations structurelles

Le modèle à équations structurelles (SEM pour « structural equation model ») encore connu sous le nom de « modèle d'analyse de structure de corrélation » est une technique statistique de modélisation des moments d'ordre 1 et 2 de données multivariées Gaussiennes [Duncan, 1975; Jöreskog, 1973; Cheung, 2015]. Le modèle général s'écrit :

$$\begin{cases} \mathbf{\Lambda} = \boldsymbol{\mu} + \mathbf{B}\mathbf{\Lambda} + \mathbf{\Gamma}\mathbf{X} + \boldsymbol{\xi} \\ \mathbf{Y} = \boldsymbol{\mu}_y + \mathbf{L}\mathbf{\Lambda} + \mathbf{K}\mathbf{X} + \boldsymbol{\epsilon} \end{cases} \quad (2.29)$$

avec $\mathbf{\Lambda}$ un D -vecteur de variables latentes, $\boldsymbol{\mu}$ un D -vecteur d'intercepts, \mathbf{B} une matrice $D \times D$ avec des zéros sur la diagonale, $\boldsymbol{\xi}$ un D -vecteur d'erreurs résiduelles centrées et indépendantes, $\mathbf{\Gamma}$ une matrice $D \times Q$ associée au Q -vecteur de variables explicatives \mathbf{X} . \mathbf{Y} un P -vecteur de variables observées, \mathbf{L} une matrice $P \times D$ appelée « factor loadings », \mathbf{K} une matrice $P \times Q$ associée au Q -vecteur de variables explicatives \mathbf{X} . $\boldsymbol{\mu}_y$ est un P -vecteur et $\boldsymbol{\epsilon}$ est un P -vecteur d'erreurs de mesure centrées indépendantes.

L'idée de l'approche par équations structurelles est d'encapsuler les hypothèses relatives à la génération des données (hypothèses causales représentées dans le DAG) dans des équations. Ces équations structurent les espérances et la matrice de variances-covariances des variables du modèle dans le but de :

- décomposer la variance de chaque variable en part expliquée par une ou plusieurs autres variables et en part non expliquée,
- et de convertir les corrélations entre les variables en paramètres de causalité (« path diagram »).

Nous allons illustrer cette idée à partir de l'exemple de la Figure 2.5 qui reprend l'exemple du DAG de la Figure 2.1. Cet exemple modélise un système $\mathbf{\Lambda}$ de trois composantes Λ_1, Λ_2 et Λ_3 . Toutes les composantes du système sont stochastiques avec un intercept μ et une part aléatoire ξ : $\Lambda_1 = \mu_1 + \xi_1$, $\Lambda_2 = \mu_2 + \xi_2$, $\Lambda_3 = \mu_3 + \xi_3$. Les variables $Y_1 = \Lambda_1 + \epsilon_1$, $Y_2 = \Lambda_2 + \epsilon_2$ et $Y_3 = \Lambda_3 + \epsilon_3$ sont les marqueurs respectifs des composantes Λ_1, Λ_2 et Λ_3 .

Le schéma causal représenté dans le DAG de la Figure 2.5 se traduit en équations structurelles suivantes :

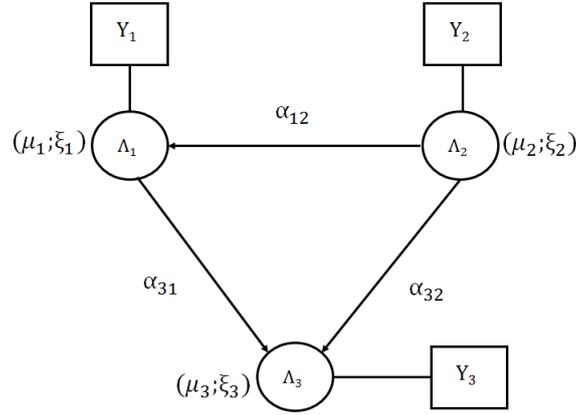


FIGURE 2.5 – DAG représentant un modèle à équations structurelles pour un système Λ de trois composantes latentes Λ_1 , Λ_2 et Λ_3 mesurées respectivement par les variables Y_1 , Y_2 et Y_3 . α_{21} , α_{31} et α_{32} modélisent les effets causaux entre les composantes du système Λ .

$$\begin{cases} \Lambda_1 = \mu_1 + \xi_1 + \alpha_{12}\Lambda_2 \\ \Lambda_2 = \mu_2 + \xi_2 \\ \Lambda_3 = \mu_3 + \xi_3 + \alpha_{31}\Lambda_1 + \alpha_{32}\Lambda_2 \\ Y_1 = \Lambda_1 + \epsilon_1 \\ Y_2 = \Lambda_2 + \epsilon_2 \\ Y_3 = \Lambda_3 + \epsilon_3 \end{cases} \quad (2.30)$$

où ξ_1 , ξ_2 , ξ_3 sont des erreurs résiduelles Gaussiennes centrées indépendantes entre elles et indépendantes des erreurs de mesure Gaussiennes centrées ϵ_1 , ϵ_2 et ϵ_3 indépendantes entre elles.

Le vecteur $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2, \Lambda_3)^\top$ est un vecteur Gaussien d'espérance $\boldsymbol{\mu}_\Lambda$ et de matrice de variance-covariance \mathbf{V}_Λ : $\mathbf{\Lambda} \sim \mathcal{N}(\boldsymbol{\mu}_\Lambda, \mathbf{V}_\Lambda)$ avec

$$\boldsymbol{\mu}_\Lambda = \begin{pmatrix} \mu_{\Lambda_1} = \mu_1 + \alpha_{12}\mu_2 \\ \mu_{\Lambda_2} = \mu_2 \\ \mu_{\Lambda_3} = \alpha_{31}\mu_1 + \alpha_{32}\mu_2 + \mu_3 \end{pmatrix}$$

$$\mathbf{V}_\Lambda = \begin{bmatrix} \sigma_{\xi_1}^2 + \alpha_{12}^2 \text{Var}(\Lambda_2) & \alpha_{12} \text{Var}(\Lambda_2) & \alpha_{31} \sigma_{\xi_1}^2 + (\alpha_{12}\alpha_{32} + \alpha_{12}^2\alpha_{31}) \text{Var}(\Lambda_2) \\ \alpha_{12} \text{Var}(\Lambda_2) & \text{Var}(\Lambda_2) & (\alpha_{31}\alpha_{12} + \alpha_{32}) \text{Var}(\Lambda_2) \\ \alpha_{31} \sigma_{\xi_1}^2 + (\alpha_{12}\alpha_{32} + \alpha_{12}^2\alpha_{31}) \text{Var}(\Lambda_2) & (\alpha_{31}\alpha_{12} + \alpha_{32}) \text{Var}(\Lambda_2) & (\sigma_{\xi_3}^2 + \alpha_{31} \sigma_{\xi_1}^2 + (\alpha_{31}\alpha_{12} + \alpha_{32})^2 \text{Var}(\Lambda_2)) \end{bmatrix}$$

avec $\text{Var}(\Lambda_2) = \sigma_{\xi_2}^2$.

La matrice de variance-covariance des variables est donc structurée par les hypothèses formulées dans les équations (2.30). Ce modèle fait l'hypothèse que la variance de Λ_2 n'est

pas expliquée ni par Λ_1 ni par Λ_3 et que les variances de ces derniers sont expliquées par Λ_2 .

Par rapport aux autres approches vues précédemment, un intérêt principal du modèle à équations structurelles est qu'il permet d'inférer des connexions entre des variables latentes et des variables observées d'un système en distinguant un modèle structurel pour le vecteur de variables latentes $\mathbf{\Lambda}$ et un modèle d'observation pour le vecteur de variables observées \mathbf{Y} [Jöreskog, 1970].

2.2.3.6 Modèle dynamique à équations structurelles

Le modèle dynamique à équations structurelles (DSEM) étend le modèle à équations structurelles aux données répétées [Muthén, 1997]. Dunson [2003] a proposé un modèle dynamique d'équations structurelles pour les données répétées de distributions de la famille exponentielle (marqueurs binaires, catégoriels, ordinaux, continus et de comptage) pour prendre en compte les corrélations entre les mesures répétées observées d'un sujet. Dunson [2003] a spécifié un modèle structurel pour des variables latentes $\mathbf{\Lambda}_{it}$ et un modèle d'observation pour le vecteur de paramètres canoniques de chaque marqueur $\boldsymbol{\theta}_{itk}$. La distribution du marqueur est ensuite établie suivant le type du marqueur conditionnellement aux paramètres canoniques du marqueur. Le modèle de Dunson [2003] s'écrit :

$$\begin{cases} \mathbf{\Lambda}_{it} = \mathbf{\Gamma}_{it}\mathbf{X}_{it} + \sum_{j=0}^{t-1} \mathbf{B}_{ij} (\mathbf{\Lambda}_{ij} - \mathbf{\Gamma}_{ij}\mathbf{X}_{ij}) + \boldsymbol{\xi}_{it} \\ h(\boldsymbol{\theta}_{itk}) = \mathbf{Z}_{itk}^\top \boldsymbol{\gamma}_k + \mathbf{\Lambda}_{it}^\top \mathbf{L}_{tk} + \zeta_{ik}, \end{cases} \quad (2.31)$$

avec $\mathbf{\Lambda}_{ij}$ un vecteur de D variables latentes au temps j , $\mathbf{\Gamma}_{ij}$ une matrice $D \times Q$ associée au vecteur de Q variables explicatives \mathbf{X}_{ij} , \mathbf{B}_{ij} une matrice $D \times D$, $\boldsymbol{\xi}_{it}$ un vecteur de taille D d'erreurs résiduelles centrées et indépendantes, h une fonction de lien, $\boldsymbol{\theta}_{itk}$ le vecteur de paramètres canoniques de la distribution du marqueur Y_{ijk} , $\boldsymbol{\gamma}_k$ un vecteur de taille P associé au vecteur de P variables explicatives \mathbf{Z}_{ijk} , \mathbf{L}_{tk} un vecteur de taille D appelé « factor loadings », ζ_{ik} un intercept aléatoire spécifique au sujet i et au marqueur Y_{ijk} .

Dunson [2003] a appliqué le modèle DSEM aux données de neurotoxicologie pour évaluer les effets d'expositions de rats au pesticide méthoxychlore (MXC) sur le comportement neurologique de leurs nourrissons. Dans cette application, une seule variable latente (le comportement neurologique) était considérée et mesurée par un test psychométrique composé de 30 items.

Récemment, Asparouhov et al. [2017] ont aussi proposé un modèle dynamique à équations structurelles pour modéliser les relations entre les variables latentes et les marqueurs répétés Gaussiens. Dans leur approche de modélisation très générale, des effets aléatoires sont introduits dans le modèle structurel des variables latentes pour tenir compte de la corrélation entre les variables latentes à différents temps pour un même sujet. De plus, les variables latentes et les variables observées sont régressées les unes sur les autres. Ce modèle s'écrit :

$$\left\{ \begin{array}{l} \Lambda_{it} = \mu_{it} + \sum_{j=0}^{t-1} \mathbf{B}_{ij}(\Lambda_{ij} - \mu_{ij} - \mathbf{Z}_{ij}\zeta_i) + \sum_{j=0}^{t-1} \mathbf{C}_{ij}(\mathbf{Y}_{ij} - \mu_{y,it}) \\ \quad + \sum_{j=0}^t \mathbf{\Gamma}_{ij}\mathbf{X}_{ij} + \mathbf{Z}_{it}\zeta_i + \xi_{it} \\ \mathbf{Y}_{it} = \mu_{y,it} + \sum_{j=0}^t \mathbf{L}_{ij}\Lambda_{ij} + \sum_{j=0}^{t-1} \mathbf{R}_{ij}(\mathbf{Y}_{ij} - \mu_{y,it}) + \sum_{j=0}^t \mathbf{K}_{ij}\mathbf{X}_{ij} + \epsilon_{it}, \end{array} \right. \quad (2.32)$$

avec :

- Λ_{ij} un D -vecteur de variables latentes au temps j ,
- μ_{ij} un vecteur de D intercepts au temps j ,
- \mathbf{B}_{ij} une matrice $D \times D$, $\zeta_i = (\zeta_{i1}^\top \dots, \zeta_{iD}^\top)^\top$ le vecteur de taille q des vecteurs d'effets aléatoires centrés Gaussiens ζ_{id} sur la composante $d = 1, \dots, D$, \mathbf{Z}_{it} la matrice $D \times q$ de variables explicatives (incluant le temps) associée aux effets aléatoires ζ_i ,
- \mathbf{C}_{ij} une matrice $D \times P$, \mathbf{Y}_{ij} un vecteur de P variables observées, $\mu_{y,it}$ un vecteur de P paramètres,
- $\mathbf{\Gamma}_{ij}$ une matrice $D \times Q$ associée au Q -vecteur de variables explicatives \mathbf{X}_{ij} .
- ξ_{it} un D -vecteur d'erreurs résiduelles centrées et indépendantes,
- \mathbf{L}_{ij} une matrice $P \times D$ appelée « factor loadings »
- \mathbf{R}_{ij} une matrice $P \times P$ de coefficients de régression,
- \mathbf{K}_{ij} une matrice $P \times Q$ associée au Q -vecteur de variables explicatives \mathbf{X}_{ij}
- ϵ_{it} un P -vecteur d'erreurs de mesure centrées indépendantes.

Ce modèle, de par sa formulation qui se veut très générale, reste compliqué à appréhender et n'est estimable que sous des simplifications et conditions d'identifiabilité précises.

Estimation du modèle à équations structurelles

La méthode du maximum de vraisemblance des observations est souvent utilisée pour l'estimation des paramètres du modèle à équations structurelles Byrne [2013]. L'estimation des paramètres du modèle à équations structurelles peut se faire dans l'approche

Bayésienne en utilisant l'algorithme MCMC combiné à l'échantillonneur de Gibbs [Muthén, 1997], notamment dans les cas plus complexes des modèles dynamiques à équations structurelles [Asparouhov et al., 2017; Dunson, 2003].

2.2.3.7 Limites de l'approche de causalité de Granger

L'inconvénient principal de l'approche de causalité de Granger et des méthodes statistiques qui en découlent vient de sa nature prédictive. Dans un système $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2, \Lambda_3)$, l'amélioration de la prédiction de la composante Λ_1 par la prise en compte de la composante Λ_2 peut être expliquée par une troisième composante Λ_3 (facteur de confusion) liée aux deux composantes Λ_1 et Λ_2 qui n'est pas prise en compte. La capacité d'une variable à prédire une autre variable mesure donc une association brute entre les deux variables qui peut avoir un sens causal.

Pearl [2003] a introduit l'outil « *do* » pour distinguer l'effet causal d'une composante Λ_2 sur une composante Λ_1 de l'association brute entre Λ_1 et Λ_2 . Pour deux composantes Λ_1 et Λ_2 d'un système $\mathbf{\Lambda}$, l'association brute entre Λ_2 et Λ_1 est quantifiée par $\mathbb{P}[\Lambda_1|\Lambda_2]$ et l'effet causal de Λ_2 sur Λ_1 par $\mathbb{P}[\Lambda_1|do(\Lambda_2)]$ [Pearl, 2003]. L'effet causal de $\Lambda_2 = \lambda_2$ par rapport à $\Lambda_2 = \lambda_2^*$ est donné par $\mathbb{P}[\Lambda_1|do(\Lambda_2 = \lambda_2)] - \mathbb{P}[\Lambda_1|do(\Lambda_2 = \lambda_2^*)]$. L'outil « *do* » s'interprète comme une intervention utopique sur la composante Λ_2 qui consisterait à la fixer à une valeur donnée sans modifier la structure de dépendance entre les composantes du système $\mathbf{\Lambda}$. La conséquence de cette intervention est que la composante Λ_2 devient fixe et la distribution de probabilité tronquée $\mathbb{P}[\mathbf{\Lambda}|do(\Lambda_2)]$ permet de mesurer l'effet que l'intervention sur la composante Λ_2 produit sur le système, et en particulier $\mathbb{P}[\Lambda_1|do(\Lambda_2)]$ mesure l'effet causal de Λ_2 sur Λ_1 .

Une intervention est naturelle dans les essais randomisés où les sujets sont aléatoirement répartis entre les modalités de la variable Λ_2 lorsque l'on s'intéresse à l'effet de cette dernière sur la variable Λ_1 . Mais dans les études en population, une intervention est utopique et l'effet causal $\mathbb{P}[\Lambda_1|do(\Lambda_2)]$ de Λ_2 sur Λ_1 doit être déduit de l'association brute observée $\mathbb{P}[\Lambda_1|\Lambda_2]$.

Facteurs de confusion et leur contrôle dans l'approche de causalité de Granger

L'approche prédictive de la causalité ne garantit pas que l'association entre deux composantes Λ_2 et Λ_1 d'un système $\mathbf{\Lambda}$ ait une interprétation causale du fait de la présence

possible de facteur de confusion. Une variable de confusion (ou facteur de confusion) est définie comme une variable qui est à la fois associée aux variables Λ_1 et Λ_2 et qui pourrait éventuellement expliquer entièrement l'association entre Λ_1 et Λ_2 [Arjas et Parner, 2004].

Il existe une riche littérature sur la réduction de biais dans l'estimation d'effet causal en présence de variable de confusion [Shrier et Platt, 2008; Greenland, 2003; Hernán et al., 2004; Pearl, 1995]. Pearl [1995] a proposé deux critères graphiques (critères de « back door » et « front-door ») pour identifier les variables d'ajustement suffisantes pour que l'association brute observée $\mathbb{P}[\Lambda_1|\Lambda_2]$ soit une estimation non biaisée de l'effet causal $\mathbb{P}[\Lambda_1|do(\Lambda_2)]$.

Ces critères sont basés sur la notion de « d-séparation » . La « d-séparation » est une généralisation de la condition de Markov à trois partitions $\{\Lambda_1\}, \{\Lambda_2\}, \{\mathbf{Z}\} \subset \{\Lambda_3, \dots, \Lambda_D\}$ d'un système Λ de D composantes. Les composantes Λ_1 et Λ_2 sont d-séparées par un ensemble de composantes $\{\mathbf{Z}\}$ si, conditionnellement à \mathbf{Z} , les variables Λ_1 et Λ_2 sont indépendantes :

$$\Lambda_1 \perp\!\!\!\perp \Lambda_2 | \mathbf{Z}.$$

Lorsque l'ensemble \mathbf{Z} d-sépare les composantes Λ_1 et Λ_2 dans chaque groupe d'individu de modalité $\mathbf{Z} = \mathbf{z}$, l'association brute $\mathbb{P}[\Lambda_1|\Lambda_2 = \lambda_2, \mathbf{Z} = \mathbf{z}]$ est une estimation non biaisée de l'effet causal $\mathbb{P}[\Lambda_1|do(\Lambda_2 = \lambda_2), \mathbf{Z} = \mathbf{z}]$:

$$\mathbb{P}[\Lambda_1|\Lambda_2 = \lambda_2, \mathbf{Z} = \mathbf{z}] = \mathbb{P}[\Lambda_1|do(\Lambda_2 = \lambda_2), \mathbf{Z} = \mathbf{z}]$$

et l'effet causal marginal non biaisé de la composante Λ_2 sur la composante Λ_1 peut se déduire par

$$\begin{aligned} \mathbb{P}[\Lambda_1|do(\Lambda_2 = \lambda_2)] &= \sum_{\mathbf{z}} \mathbb{P}[\Lambda_1 = \lambda_1 | \Lambda_2 = \lambda_2, \mathbf{Z} = \mathbf{z}] \mathbb{P}[\mathbf{Z} = \mathbf{z}] \\ &= \sum_{\mathbf{z}} \mathbb{P}[\Lambda_1 = \lambda_1, \mathbf{Z} = \mathbf{z} | \Lambda_2 = \lambda_2] \\ &= \mathbb{P}[\Lambda_1 | \Lambda_2 = \lambda_2] \end{aligned} \tag{2.33}$$

2.2.4 Approche mécaniste de la causalité

Une troisième approche de la causalité existe pour appréhender des systèmes dynamiques en temps continu. Cette approche qui étend la causalité de Granger [1969] aux systèmes dynamiques en temps continu repose sur les travaux de Mykland [1986] et plus récemment de Petrović [1996]; Commenges et Gégout-Petit [2009]. Conceptuellement, cette approche s'appuie sur des graphes de dépendance locale (LDG comme illustré en Figure

2.3) dans lesquels l'influence causale du processus Λ_d sur le processus $\Lambda_{d'}$ est quantifiée par l'effet du processus Λ_d sur le changement instantané du processus $\Lambda_{d'}$. Ces influences instantanées sont naturellement traduites en modèle statistique à base de système d'équations différentielles.

Le modèle à équations différentielles est classique pour inférer des relations causales dans un système de composantes identifiées à des processus stochastiques évoluant en temps continu.

La causalité dans les modèles d'équations différentielles est formulée à l'aide de notions théoriques de filtrations de processus stochastiques, de système autonome, d'orthogonalité de martingales et de décomposition de Doob-Meyer sur lesquelles nous revenons brièvement.

2.2.4.1 Causalité, notions de filtration et de système autonome

La notion de causalité mécaniste ou stochastique repose sur la notion de filtration et est formulée par [Mykland, 1986] comme suit.

Dans un espace filtré de probabilité $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$, un processus stochastique $(\mathbf{\Lambda}_{1t})_{t \geq 0}$ de filtration $(\mathcal{F}_t^{\mathbf{\Lambda}_1})$ « cause entièrement » un autre processus $(\mathbf{\Lambda}_{2t})_{t \geq 0}$ de filtration $(\mathcal{F}_t^{\mathbf{\Lambda}_2})$ dans la filtration (\mathcal{F}_t) si

- les filtrations $(\mathcal{F}_t^{\mathbf{\Lambda}_1})$ et $(\mathcal{F}_t^{\mathbf{\Lambda}_2})$ sont incluses dans la filtration $(\mathcal{F}_t)_{t \geq 0}$
- pour tout évènement $A \in (\mathcal{F}_\infty^{\mathbf{\Lambda}_2}), \mathbb{P}[A | (\mathcal{F}_t)_{t \geq 0}] = \mathbb{P}[A | (\mathcal{F}_t^{\mathbf{\Lambda}_1})_{t \geq 0}]$

La notion de causalité stochastique telle que définie par Mykland [1986] est intuitive pour étudier les relations causales dans un univers U à partir de l'accroissement d'informations sur U au cours du temps t , c'est-à-dire $(\mathcal{F}_t)_{t \geq 0}$. L'univers U contient les composantes (ou systèmes) dynamiques $\mathbf{\Lambda}_1$ et $\mathbf{\Lambda}_2$ identifiées par les processus stochastiques $(\mathbf{\Lambda}_{1t})_{t \geq 0}$ et $(\mathbf{\Lambda}_{2t})_{t \geq 0}$. La composante (ou le système) $\mathbf{\Lambda}_1$ « cause entièrement » la composante (ou le système) $\mathbf{\Lambda}_2$ dans l'univers U si les informations disponibles sur $\mathbf{\Lambda}_1$ (c'est-à-dire la filtration $(\mathcal{F}_t^{\mathbf{\Lambda}_1})$) sont suffisantes pour complètement prédire la dynamique de $\mathbf{\Lambda}_2$. Autrement dit, à un temps t quelconque, la quantité d'informations sur $\mathbf{\Lambda}_2$ apportées par la filtration de tout l'univers U n'est pas plus grande que la quantité d'informations apportées par la filtration de $\mathbf{\Lambda}_1$.

Notion de système autonome

Lorsque dans un univers U , les informations disponibles sur un système $\mathbf{\Lambda}$ sont suffi-

santes pour complètement prédire sa propre dynamique, le système Λ est sa propre cause [Petrović, 1996]. Le système Λ est alors dit autonome ou « suffisamment riche » [Commenges et Gégout-Petit, 2009].

Pour explorer les relations causales dans un système Λ_t , l'hypothèse de système « suffisamment riche » est faite. Il est important de faire cette hypothèse pour exclure les influences externes au système.

Absence de causalité

Il n'existe pas de relation causale entre deux systèmes Λ_1 et Λ_2 si et seulement si les deux systèmes sont autonomes dans l'univers U déterminé par les deux systèmes Λ_1 et Λ_2 .

Réciproquement si toutes les composantes d'un système $\Lambda = (\Lambda_1, \dots, \Lambda_D)$ sont autonomes alors il n'existe pas de relation causale entre les composantes du système.

2.2.4.2 Décomposition de Doob-Meyer et notion d'orthogonalité

Considérons un système autonome $\Lambda = (\Lambda_1, \dots, \Lambda_D)$ identifié par les processus stochastiques $(\Lambda_d(t))_{t \geq 0, d=1, D}$ appartenant à la famille des processus semi-martingales.

Décomposition de Doob-Meyer

Tout processus semi-martingale $\Lambda_d(t)_{t \geq 0}$ admet la décomposition de Doob-Meyer suivant :

$$\Lambda_d(t) = \Gamma_d(t) + M_d(t) \quad (2.34)$$

où $(\Gamma_d(t))_{t \geq 0}$ est un processus déterministe appelé compensateur de $(\Lambda_d(t))_{t \geq 0}$ et $(M_d(t))_{t \geq 0}$ est une martingale.

De manière analogue, la décomposition de Doob-Meyer du processus multivarié $(\Lambda(t))_{t \geq 0} = (\Lambda_d(t))_{t \geq 0, d=1, D}$ s'écrit :

$$\Lambda(t) = \Gamma(t) + \mathbf{M}(t) \quad (2.35)$$

avec $\Gamma(t)_{t \geq 0} = (\Gamma_d(t))_{t \geq 0, d=1, D}$ le compensateur de $(\Lambda(t))_{t \geq 0}$ et $\mathbf{M}(t) = (M_d(t))_{t \geq 0, d=1, D}$ est une martingale multivariée de D composantes.

Le but de la décomposition de Doob-Meyer du processus multivarié $(\Lambda(t))_{t \geq 0}$ en processus déterministes $\Gamma(t)_{t \geq 0}$ et en martingale multivariée $\mathbf{M}(t) = (M_d(t))_{t \geq 0, d=1, D}$ est d'inférer la structure de causalité dans le système $\Lambda = (\Lambda_1, \dots, \Lambda_D)$ à travers la modélisation conjointe des processus déterministes $(\Gamma_d(t))_{t \geq 0, d=1, D}$.

Cela revient à faire l'hypothèse de non-existence de relations causales entre les composantes du martingale $\mathbf{M}(t) = ((M_d(t))_{t \geq 0, d=1, D})$.

Martingales et notion d'orthogonalité

Sur la base de la formulation de la causalité de Mykland [1986], l'inexistence de relations causales entre des martingales est équivalente à l'orthogonalité des martingales [Petrović, 1996].

À partir de la décomposition de Doob-Meyer du processus $(\mathbf{\Lambda}(t))_{t \geq 0} = (\Lambda_d(t))_{t \geq 0, d=1, D}$ et sous l'hypothèse que les martingales $(M_d(t))_{d=1, D}$ sont orthogonales, un processus $(\Lambda_d(t))_{t \geq 0} \in \mathbf{\Lambda}(t)_{t \geq 0}$ influence un processus $(\Lambda_{d'}(t))_{t \geq 0} \in \mathbf{\Lambda}(t)_{t \geq 0}$ si le compensateur du processus $(\Lambda_{d'}(t))_{t \geq 0}$ dépend des valeurs passées du processus $(\Lambda_d(t))_{t \geq 0}$ [Florens et Fougère, 1996; Commenges et Gégout-Petit, 2009].

De plus, si le compensateur $\mathbf{\Gamma}(t) = (\Gamma_1(t), \dots, \Gamma_D(t))$ est différentiable alors la décomposition de Doob-Meyer peut être formulée sous forme d'équations différentielles stochastiques :

$$d\mathbf{\Lambda}(t) = \boldsymbol{\gamma}(t)dt + d\mathbf{M}(t), \quad (2.36)$$

avec $\boldsymbol{\gamma}(t) = \frac{d\mathbf{\Gamma}(t)}{dt}$ est la dérivée par rapport au temps du processus déterministe $\mathbf{\Gamma}(t)$ et $\mathbf{M}(t)$ est une martingale multivariée de D composantes.

Soulignons que ce modèle repose sur deux hypothèses fondamentales que sont :

- le système $\mathbf{\Lambda}$ est identifié à un processus stochastique multivarié $(\mathbf{\Lambda}(t))_{t \geq 0} = (\Lambda_d(t))_{t \geq 0, d=1, D}$ qui est autonome au sens de Mykland [1986],
- $(\mathbf{\Lambda}(t))_{t \geq 0} = (\Lambda_d(t))_{t \geq 0, d=1, D}$ est une semi-martingale qui admet une décomposition de Doob-Meyer dans laquelle les martingales sont orthogonales.

2.2.4.3 Modèle statistique à équations différentielles

Dans un échantillon de n sujets $i = 1, \dots, N$, on considère un système dynamique $\mathbf{\Lambda}_i = (\Lambda_{i1}, \dots, \Lambda_{iD})$ mesuré à partir de K marqueurs longitudinaux $\mathbf{Y}_{ij} = (Y_{ij1}, \dots, Y_{ijK})$ chez le sujet i aux temps d'observation t_{ijk} . L'inférence causale par des équations différentielles stochastiques consiste à définir un modèle d'équations différentielles stochastiques pour les composantes du système $\mathbf{\Lambda}_i$ et un modèle d'observation qui relie les K marqueurs au système. Le modèle statistique s'écrit :

$$\begin{cases} d\mathbf{\Lambda}_i(t) &= \boldsymbol{\gamma}_i(t)dt + d\mathbf{M}_i(t) \\ Y_{ijk} &= f_k(\mathbf{\Lambda}_i(t_{ijk})) + \epsilon_{ijk}, \end{cases} \quad (2.37)$$

avec $\gamma_i(t)$ la dérivée du compensateur du processus Λ_i , $\mathbf{M}_i(t)$ une martingale multivariée de D composantes orthogonales, f_k une fonction de lien et ϵ_{ijk} des erreurs de mesure indépendantes de variance σ_k^2 .

La dynamique des composantes du système Λ_i et leurs relations causales peuvent être modélisées par l'effet de variables explicatives sur $\gamma_i(t)$. Avec un mouvement Brownien comme martingale \mathbf{W}_i dans le modèle 2.37, ce dernier s'écrit :

$$\begin{cases} \frac{d\Lambda_i(t)}{dt} &= \mathbf{x}_i(t)\boldsymbol{\beta} + d\mathbf{W}_i(t) \\ Y_{ijk} &= f_k(\Lambda_i(t_{ijk})) + \epsilon_{ijk}, \end{cases} \quad (2.38)$$

avec $\boldsymbol{\beta}$ les effets associés aux variables dépendantes du temps $\mathbf{x}_i(t)$ et $\mathbf{W}_i(t)$ un mouvement Brownien.

Les relations entre les marqueurs et les composantes du système peuvent aussi être modélisées par des effets de variables explicatives. Ainsi, le modèle 2.38 peut s'écrire :

$$\begin{cases} \frac{d\Lambda_i(t)}{dt} &= \mathbf{x}_i(t)\boldsymbol{\beta} + d\mathbf{W}_i(t) \\ Y_{ijk} &= f_k(\Lambda_i(t_{ijk}), \mathbf{z}_{ik}^\top \boldsymbol{\alpha}_k) + \epsilon_{ijk}, \end{cases} \quad (2.39)$$

avec $\boldsymbol{\alpha}_k$ des effets de variables explicatives \mathbf{z}_{ik} spécifiques au marqueur k . Notons que dans ces modèles, certains paramètres peuvent être aléatoires.

Dans le cas où le système Λ_i est déterministe, le modèle statistique devient un modèle à équations différentielles ordinaires (ODE, pour Ordinary Differential Equations). Les modèles statistiques à base d'ODE sont beaucoup plus courants en biostatistique car ils sont moins complexes à estimer que ceux à base d'équations différentielles stochastiques.

2.2.4.4 Estimation du modèle à équations différentielles

Avant leur utilisation en biostatistique, les équations différentielles ont été utilisées pour prédire des trajectoires de marqueurs selon des valeurs théoriques des paramètres (par exemple [Ehrenstein et al., 1997]). En biostatistique, les modèles à équations différentielles ont été utilisés pour modéliser des mécanismes de phénomènes bien connus théoriquement. Ces modèles ont d'abord été estimés à l'échelle individuelle. Plus récemment, des méthodes d'estimation des paramètres du modèle à base de ODE à partir de données de population ont été proposées.

Dans le cas où il existe une solution analytique aux équations différentielles, l'approche classique d'estimation par maximum de vraisemblance est utilisée. Mais souvent il n'existe

pas de solution analytique et l'estimation des paramètres du modèle à équations différentielles est compliquée par la résolution numérique des équations différentielles à l'aide de solveur. Cela induit un temps de calcul extrêmement important.

Dans le cas d'équations différentielles ordinaires (ODE), plusieurs approches d'inférence ont été proposées. Il y a l'approche paramétrique qui est basée sur l'estimateur du maximum de vraisemblance [Guedj et al., 2007] ou l'estimateur du maximum de vraisemblance pénalisée [Prague et al., 2016]. Les estimateurs du maximum de vraisemblance peuvent être obtenus en maximisant directement la vraisemblance des données observées ou leur vraisemblance pénalisée par des algorithmes de type Newton-Raphson. Cette approche est implémentée par exemple dans le logiciel NIMROD [Prague et al., 2013].

Les estimateurs du maximum de vraisemblance peuvent aussi être obtenus en maximisant la vraisemblance complète des données par un algorithme EM ou sa version stochastique SAEM (pour « Stochastic Approximation of EM ») implémentée par exemple dans le programme MONOLIX [Monolix]. Dans ces algorithmes, les composantes non observées dans les équations différentielles sont explicitement prises en compte dans le calcul de la vraisemblance complète qui est maximisée en deux étapes. Dans l'étape E, les composantes non observées sont prédites sachant les données observées et dans l'étape M la vraisemblance des données complètes est maximisée. Dans la version SAEM, les valeurs prédites des composantes non observées sont obtenues à partir d'une chaîne de Markov qui converge en loi vers la distribution des composantes non observées conditionnellement aux données observées.

D'autres approches d'estimation non-paramétriques basées sur une linéarisation des équations différentielles ont été aussi proposées [Liang et Wu, 2008; Ramsay et al., 2007].

2.2.4.5 Applications de l'approche mécaniste de causalité

En épidémiologie, l'approche de causalité par équations différentielles a été très utilisée pour comprendre le mécanisme du VIH. Prague et al. [2016] ont par exemple utilisé le modèle à équations différentielles ordinaires pour inférer les dynamiques du virus (V) et des cellules CD4 en distinguant celles qui sont au repos (Q) de celles qui sont actives et non-infectées (T) ou actives et infectées (T^*). Leur modèle s'écrit :

$$\begin{cases} \frac{dQ_t^i}{dt} &= \lambda^i + \rho^i T_t^i - \alpha_i Q_t^i - \mu_Q^i Q_t^i, \\ \frac{dT_t^i}{dt} &= \alpha^i Q_t^i - \gamma^i T_t^i - \rho^i T_t^i - \mu_T^i T_t^i, \\ \frac{dT_t^{*i}}{dt} &= \gamma^i T_t^i V_t^i - \mu_{T^*}^{*i} T_t^{*i}, \\ \frac{dV_t^i}{dt} &= \pi^i T_t^{*i} - \mu_V^i V_t^i. \end{cases} \quad (2.40)$$

ÉTAT DE L'ART

avec $\lambda^i, \rho^i, \alpha_i, \mu_Q^i, \alpha^i, \gamma^i, \rho^i, \mu_T^i, \mu_{T^*}^i, \pi^i, \mu_V^i$ les paramètres du modèle spécifiques au sujet i .

Un modèle d'observation est adossé au système d'équations différentielles pour lier les quantités V, Q, T et T^* aux observations du nombre total de cellules CD4 Y_j^i et de la charge virale VL_j^i mesurés au temps d'observation t_{ij} . Le modèle d'observation s'écrit :

$$\begin{cases} (Y_j^i)^{1/4} &= (Q_{t_{ij}}^i + T_{t_{ij}}^i + T_{t_{ij}}^{*i})^{1/4} + \epsilon_{ij}^1 \\ \log_{10} VL_j^i &= \log_{10} V_{t_{ij}}^i + \epsilon_{ij}^2. \end{cases} \quad (2.41)$$

avec ϵ_{ij}^1 et ϵ_{ij}^2 des erreurs de mesure Gaussiennes centrées et indépendantes.

Dans ce travail, l'effet d'un traitement A_t^i de type HART (pour « High active antiretroviral therapy ») est évalué sur la dynamique du VIH à travers le paramètre d'infectivité γ^i qui est regressé sur A_t^i . Ce modèle a donc permis d'inférer l'effet du traitement sur le nombre de cellules CD4 et la charge virale en tenant compte du mécanisme biologique bien connu du VIH qui lie les cellules CD4 (inactifs, infectés, non-infectés) à la charge virale.

Dans l'étude de la maladie d'Alzheimer, l'approche de causalité par équations différentielles est quasi inexistante du fait du peu de connaissances sur le mécanisme biologique de la maladie d'Alzheimer. Toutefois, Ehrenstein et al. [1997] a utilisé l'approche de causalité par équations différentielles dans un modèle hypothétique de la relation dynamique entre la perte d'acétylcholine, un marqueur de la neuro-dégénérescence, et l'accumulation de la protéine β -amyloïde. Dans ce travail, il n'était pas question d'estimation de paramètres d'un modèle à équations différentielles sur une population mais uniquement de décrire par un modèle mathématique les trajectoires de perte d'acétylcholine en fonction de différentes valeurs pré-définies du taux d'accroissement de β -amyloïde. Cela étant, c'est à notre connaissance le seul modèle à équations différentielles en lien avec la maladie d'Alzheimer décrit dans la littérature.

2.2.4.6 Avantages et limites de l'approche mécaniste de causalité

L'approche stochastique de causalité est souvent utilisée pour inférer plus finement des paramètres d'un modèle décrivant la dynamique d'un système (physique ou biologique) dont le mécanisme est bien connu. C'est le cas dans le VIH ou encore dans certains cancers [Desmée et al., 2017].

Elle présente les avantages d'être en temps continu et de modéliser des relations pos-

siblement non-linéaires et dynamiques entre les composantes d'un système stochastique même si dans la pratique les composantes sont le plus souvent supposées déterministes et des équations différentielles ordinaires remplacent les équations différentielles stochastiques.

Cependant, cette approche présente quelques limites. Les principales limites sont liées au temps extrêmement long de l'estimation et au problème d'identifiabilité de ces modèles. Quelle que soit la méthode d'estimation, les modèles à équations différentielles sont le plus souvent trop coûteux en temps de calcul et trop complexes pour être estimés directement. Des simplifications sont nécessaires pour atteindre une identifiabilité numérique. Cela peut se faire par exemple en fixant des paramètres à des valeurs déjà établies dans la littérature (comme la clairance d'un type de cellule ou le taux d'apparition).

De plus, cette approche est très compliquée à mettre en oeuvre dans des échantillons de grande taille comme dans les études en population ou lorsque le système sous-jacent est mal connu et/ou fait intervenir beaucoup de composantes comme dans la maladie d'Alzheimer. Dans ces cas, des problèmes de calculs surviennent très rapidement avec la taille de l'échantillon et avec le nombre de composantes du système et la complexité des relations.

Ces limites ont largement contribué à réduire leur utilisation en pratique malgré l'intérêt de l'approche pour l'inférence causale.

3 Méthodologie

Ce chapitre est consacré aux développements théoriques proposés au cours de la thèse et à la démarche d'implémentation du modèle correspondant dans un package R.

Ce chapitre commence par la présentation du modèle structurel et du modèle d'observation proposés. Ensuite nous établissons l'expression de la log-vraisemblance et présentons l'approche de maximisation de la log-vraisemblance. Le modèle étant défini en temps discret, nous explorons ensuite l'impact de la discrétisation du temps sur l'interprétation des estimations du modèle. Enfin, nous abordons les prédictions et l'implémentation du modèle dans un package R.

3.1 Présentation du modèle

3.1.1 Modèle structurel

3.1.1.1 Présentation du modèle structurel

Considérons un réseau de processus latents $\mathbf{\Lambda}_i(t)_{t \geq 0} = (\Lambda_{id}(t))_{d=1, \dots, D}^\top$ représentant un système de D dimensions dynamiques chez un individu i avec $i = 1, \dots, N$. Nous supposons que le réseau de processus latents $\mathbf{\Lambda}_i(t)_{t \geq 0}$ est défini en $J + 1$ temps discrets $t = j \times \delta$ avec $j = 0, 1, \dots, J$, et δ un pas de discrétisation constant. Notons $\Delta \mathbf{\Lambda}_i(t + \delta) = \mathbf{\Lambda}_i(t + \delta) - \mathbf{\Lambda}_i(t)$ la variation du système entre deux temps consécutifs et $\frac{\Delta \mathbf{\Lambda}_i(t + \delta)}{\delta}$ le taux de variation du système entre deux temps consécutifs.

La dynamique du réseau de processus latents est modélisée par un modèle structurel qui combine deux sous-modèles :

- un sous-modèle pour le niveau initial du système de processus latents,
- un sous-modèle pour le taux de variation du système de processus latents au cours du temps.

Le niveau initial des processus est modélisé par un modèle linéaire multivarié à effets mixtes. Le taux de variation du système de processus est modélisé par un modèle

linéaire multivarié à effets mixtes avec une composante auto-régressive. Le modèle structurel s'écrit :

$$\begin{cases} \Lambda_i(0) = \mathbf{X}_i^{(0)}\boldsymbol{\beta} + \mathbf{u}_i \\ \frac{\Delta\Lambda_i(t+\delta)}{\delta} = \mathbf{X}_i(t+\delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t+\delta)\mathbf{v}_i + \mathbf{A}_{i,\delta}(t)\Lambda_i(t), \quad \forall t \geq 0, \end{cases} \quad (3.1)$$

avec $\mathbf{u}_i = (u_{id})_{d=1,D}$ un vecteur de taille D et u_{id} l'effet aléatoire Gaussien spécifique à l'individu i sur le niveau initial du processus d . Le vecteur $\mathbf{v}_i = (\mathbf{v}_{id}^\top)_{d=1,D}$ est de taille $q = \sum_{d=1}^D q_d$ avec \mathbf{v}_{id} un vecteur d'effets aléatoires Gaussiens spécifiques à l'individu i associés au taux de variation du processus d au cours du temps. La structure de la matrice de variance-covariance des effets aléatoires \mathbf{u}_i et \mathbf{v}_i est explicitée en section 3.1.1.3. $\mathbf{X}_i^{(0)}$ est la matrice $D \times p_0$ de variables explicatives associées au vecteur de taille p_0 des effets fixes $\boldsymbol{\beta}$, $\mathbf{X}_i(t)$ est la matrice $(D \times p)$ de variables explicatives associées au p -vecteur $\boldsymbol{\gamma}$ d'effets fixes et $\mathbf{Z}_i(t)$ est la matrice de taille $(D \times q)$ de variables explicatives associées au q -vecteur \mathbf{v}_i des effets aléatoires spécifiques à l'individu i . Enfin $\mathbf{A}_{i,\delta}(t)$ est une matrice de taille $D \times D$.

3.1.1.2 Influences temporelles entre les processus

La matrice $\mathbf{A}_{i,\delta}(t)$ de la composante auto-régressive modélise les influences temporelles entre les processus latents. Plus précisément, suivant l'approche stochastique de la causalité, $\mathbf{A}_{i,\delta}(t)$ modélise à chaque temps t l'effet des processus latents sur le taux de variation à venir du réseau, sur un intervalle de temps δ . Cette matrice d'influences temporelles est centrale dans notre approche de modélisation simultanée de plusieurs processus et de leurs inter-relations.

Au temps t la matrice $\mathbf{A}_{i,\delta}(t)$ est définie par

$$\mathbf{A}_{i,\delta}(t) = \begin{pmatrix} a_{i,11}(t) & \dots & a_{i,1d}(t) & \dots & a_{i,1D}(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,d1}(t) & \dots & a_{i,dd}(t) & \dots & a_{i,dD}(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,D1}(t) & \dots & a_{i,Dd}(t) & \dots & a_{i,DD}(t) \end{pmatrix}$$

où chaque coefficient $a_{i,dv}(t)$ de la matrice d'influences temporelles $\mathbf{A}_{i,\delta}(t)$ modélise au temps t l'influence temporelle du processus d' sur le taux de variation du processus d entre les temps t et $t + \delta$.

Pour tenir compte de la possible variation au cours du temps ou suivant des caractéristiques individuelles de la structure d'influences temporelles entre les processus, nous

MÉTHODOLOGIE

avons proposé un modèle de régression linéaire pour chaque coefficient $a_{i,dd'}(t)$:

$$a_{i,dd'}(t) = \alpha_{dd'}^0 + \sum_{k=1}^m \alpha_{dd'}^k C_{ik}(t), \quad (3.2)$$

avec $\alpha_{dd'}^0$ l'intercept et $\alpha_{dd'}^k$ l'effet de la variable explicative possiblement dépendante du temps $C_{ik}(t)$ ($k = 1, \dots, m$).

Par exemple, $C_{ik}(t)$ peut inclure des caractéristiques des sujets pour évaluer des structures d'influences temporelles différentes selon des profils de sujets. Ce sera le cas dans notre application avec des influences possiblement différentes suivant des stades cliniques de la maladie d'Alzheimer.

Les variables $C_{ik}(t)$ peuvent aussi caractériser une éventuelle évolution temporelle de la structure d'influence d'un processus sur un autre. En considérant une base de splines, l'influence dynamique du processus d' sur le processus d devient :

$$a_{i,dd'}(t) = \alpha_{dd'}^0 + \sum_{k=1}^m \alpha_{dd'}^k \mathcal{B}_k^{(l)}(t), \quad (3.3)$$

avec $\mathcal{B}_k^{(l)}$ ($k = 1, m$) une base de m splines de degré l construites à partir de n nœuds (avec $m = n - 1 + l$).

Dans la suite nous posons $\boldsymbol{\alpha}_{dd'} = (\alpha_{dd'}^k)_{k=0,m}$ le vecteur de paramètres associés à la modélisation du coefficient $a_{i,dd'}(t)$ et $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_{dd'}^\top)_{d=1,D,d'=1,D}^\top$ le vecteur de paramètres associés à la modélisation de la matrice d'influences temporelles $A_{i,\delta}(t)$.

3.1.1.3 Corrélations résiduelles entre les processus

La matrice d'influences temporelles $A_{i,\delta}$ induit une structure de corrélation entre les processus latents et entre les mesures répétées d'un même processus. Cette structure de corrélation peut être assouplie davantage par le biais des corrélations entre les effets aléatoires \mathbf{u}_i et \mathbf{v}_d .

Corrélations intra-processus

Les effets aléatoires u_{id} et v_{id} spécifiques à l'individu i et au processus d peuvent être corrélés pour capter les corrélations entre les mesures répétées du même processus d au-delà de la structure auto-régressive. Pour cela on peut supposer que les effets aléatoires u_{id} et v_{id} forment un vecteur Gaussien centré de matrice de variance-covariance

$$\begin{pmatrix} b_{ud} & \mathbf{B}_{uvd}^\top \\ \mathbf{B}_{uvd} & \mathbf{B}_{vd} \end{pmatrix}$$

avec b_{ud} la variance de l'effet aléatoire u_{id} sur le niveau initial du processus d , \mathbf{B}_{vd} la matrice $q_d \times q_d$ de variance-covariance du vecteur d'effets aléatoires \mathbf{v}_{id} sur le taux de variation du processus, et \mathbf{B}_{uvd}^\top le vecteur de taille q_d de covariance de $(u_{id}, \mathbf{v}_{id})$.

Corrélations inter-processus

Au temps $t = 0$, il peut exister des corrélations entre les niveaux des processus qui sont dues à la dynamique du système antérieure à $t = 0$. Nous avons capté ces possibles corrélations entre les niveaux initiaux des processus latents en corrélant les effets aléatoires $(u_{id})_{d=1,D}$ sur le niveau initial du système de processus latents. Le vecteur des effets aléatoires $\mathbf{u}_i = (u_{i1}, \dots, u_{iD})$ sur les niveaux des processus à $t = 0$ est donc supposé Gaussien avec la matrice $D \times D$ de variance-covariance \mathbf{B}_u non structurée.

Afin de s'assurer que la matrice d'influences temporelles $A_{i,\delta}$ modélise à elle seule les influences entre les processus latents après le temps $t = 0$, nous avons supposé que les effets aléatoires sur les taux de variations des processus sont indépendants entre processus et les effets aléatoires sur le taux de variation d'un processus d sont indépendants des effets aléatoires sur les niveaux initiaux des autres processus.

Avec ces spécifications, le vecteur $\mathbf{w}_i = \left((u_{i1}, \mathbf{v}_{i1}^\top), \dots, (u_{iD}, \mathbf{v}_{iD}^\top) \right)^\top$ de tous les effets aléatoires spécifiques à l'individu i a une distribution Gaussienne centrée :

$$\mathbf{w}_i \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{B}_u & \mathbf{B}_{uv} \\ \mathbf{B}_{uv}^\top & \mathbf{B}_v \end{pmatrix} \right)$$

avec \mathbf{B}_u la matrice non structurée de taille $D \times D$ de variance-covariance des effets aléatoires sur le niveau initial du système de processus latents, \mathbf{B}_v une matrice diagonale par bloc de taille $q \times q$ où le bloc d vaut \mathbf{B}_{vd} , et \mathbf{B}_{uv} une matrice $D \times q$ dont la $d^{\text{ième}}$ ligne vaut $\left(\mathbf{0}_{\sum_{l=1}^{d-1} q_l}, \mathbf{B}_{uvd}, \mathbf{0}_{\sum_{l=d+1}^D q_l} \right)$ où $\mathbf{0}_x$ est un x -vecteur ligne d'élément 0.

Nous notons ζ l'ensemble des paramètres de la matrice \mathbf{B} .

3.1.1.4 Exemple illustratif du modèle structurel

Pour illustrer notre approche de modélisation de processus latents, considérons un système $\Lambda_i(t) = (\Lambda_{i1}(t), \Lambda_{i2}(t))$ pour $t = 0, \delta, \dots, J\delta$. Supposons que chaque processus a un effet aléatoire sur le niveau initial : u_{1i} pour le processus Λ_{1i} et u_{2i} pour le processus Λ_{2i} ; et un effet aléatoire sur le taux de variation au cours du temps : v_{1i} pour le processus Λ_{1i} et v_{2i} pour le processus Λ_{2i} .

MÉTHODOLOGIE

Le modèle structurel pour les deux processus latents qui n'inclut pas de variable explicative sur le niveau initial et qui n'inclut pas de variable explicative sur le taux de variation des processus s'écrit :

$$\left\{ \begin{array}{l} \Lambda_{1i}(0) = \beta_1 + u_{1i} \\ \Lambda_{2i}(0) = \beta_2 + u_{2i} \\ \frac{\Lambda_{1i}(t+\delta) - \Lambda_{1i}(t)}{\delta} = \gamma_1 + v_{1i} + a_{i,11}\Lambda_{1i} + a_{i,12}\Lambda_{2i} \\ \frac{\Lambda_{2i}(t+\delta) - \Lambda_{2i}(t)}{\delta} = \gamma_2 + v_{2i} + a_{i,21}\Lambda_{1i} + a_{i,22}\Lambda_{2i} \end{array} \right.$$

Influences temporelles entre les processus

Les influences temporelles entre les deux processus sont captées par la matrice

$$\mathbf{A}_i = \begin{pmatrix} a_{i,11} & a_{i,12} \\ a_{i,21} & a_{i,22} \end{pmatrix}$$

La matrice A_i induit une structure de corrélation intra et inter processus.

Corrélations résiduelles intra-processus

La corrélation intra-processus peut être assouplie davantage en corrélant les effets aléatoires spécifiques à chaque processus :

$$(u_{1i}, v_{1i}) \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} b_{u1}^2 & b_{u1v1} \\ b_{u1v1} & b_{v1}^2 \end{pmatrix} \right)$$

et

$$(u_{2i}, v_{2i}) \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} b_{u2}^2 & b_{u2v2} \\ b_{u2v2} & b_{v2}^2 \end{pmatrix} \right)$$

Corrélations résiduelles inter-processus

La corrélation résiduelle inter-dimensions possiblement due à la dynamique des deux processus antérieure à $t = 0$ est captée en corrélant les effets aléatoires sur les niveaux initiaux des processus :

$$(u_{1i}, u_{2i}) \sim \mathcal{N} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} b_{u1}^2 & b_{u1u2} \\ b_{u1u2} & b_{u2}^2 \end{pmatrix} \right)$$

Avec ces spécifications, le vecteur de tous les effets aléatoires $\mathbf{w}_i = ((u_{1i}, v_{1i}), (u_{2i}, v_{2i}))^\top$ est Gaussien centré de matrice de variance-covariance B :

$$\begin{pmatrix} \begin{pmatrix} b_{u1}^2 & b_{u1v1} \\ b_{u1v1} & b_{v1}^2 \end{pmatrix} & \begin{pmatrix} b_{u1u2} & 0 \\ 0 & 0 \end{pmatrix} \\ \begin{pmatrix} b_{u1u2} & 0 \\ 0 & 0 \end{pmatrix} & \begin{pmatrix} b_{u2}^2 & b_{u2v2} \\ b_{u2v2} & b_{v2}^2 \end{pmatrix} \end{pmatrix}$$

3.1.2 Modèle d'observation pour données complètes

Considérons K ($K \geq D$) marqueurs longitudinaux continus $\mathbf{Y}_{ij} = (Y_{ijk})_{k=1,\dots,K}^\top$ mesurés chez l'individu i à $(n_i + 1)$ temps d'observation discrets $t_{ij} = j \times \delta$ tel que $j \in \tau_i = \{j_{0i}, \dots, (j_{0i} + n_i)\}$. Posons $\mathbf{Y}_i = (\mathbf{Y}_{ij}^\top)_{j \in \tau_i}^\top$ le vecteur de taille $K_i = (n_i + 1) \times K$ représentant les données complètes de l'individu i , et $\mathbf{Y} = (\mathbf{Y}_i^\top)_{i=1,N}^\top$ le vecteur de taille $\sum_{i=1}^N K_i$ représentant les données complètes des N individus. Nous supposons pour l'instant qu'il n'y a pas de données manquantes pour le sujet i .

Suivant la même approche de modélisation à processus latent proposée dans Proust-Lima et al. [2006, 2013], nous avons supposé que chaque processus Λ_{id} était un trait latent de l'individu i mesuré par K_d marqueurs continus ($K = \sum_{d=1}^D K_d$) et nous notons \mathcal{K}_d l'ensemble des indices des marqueurs qui mesurent le même processus d . Nous supposons ici que chaque marqueur mesure un et un seul processus latent. Dans la suite nous distinguons deux cas de figure. Le cas où les marqueurs sont Gaussiens et l'extension au cas où les marqueurs sont continus et possiblement non Gaussiens.

3.1.2.1 Modèle d'observation pour marqueur Gaussien

Le modèle d'observation consiste à supposer que le marqueur Gaussien Y_{ijk} est une mesure bruitée du processus latent d . Cela se traduit par le modèle de régression linéaire :

$$Y_{ijk} = \eta_{0k} + \eta_{1k} \Lambda_{id}(t_{ij}) + \epsilon_{ijk}, \quad \forall k \in \mathcal{K}_d \ \& \ \forall j \in \tau_i \quad (3.4)$$

qui peut se réécrire :

$$\frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} = \tilde{Y}_{ijk} = \Lambda_{id}(t_{ij}) + \epsilon_{ijk}, \quad \forall k \in \mathcal{K}_d \ \& \ \forall j \in \tau_i, \quad (3.5)$$

où le vecteur de paramètres $\boldsymbol{\eta}_k = (\eta_{0k}, \eta_{1k})^\top$ est utilisé pour construire un marqueur intermédiaire standardisé \tilde{Y}_{ijk} et ϵ_{ijk} sont des erreurs de mesure indépendantes de variance σ_k^2 .

3.1.2.2 Modèle d'observation pour marqueur continu non Gaussien

Pour prendre en compte un éventuel écart à la normalité du marqueur, nous proposons un modèle d'observation plus général. Il consiste à normaliser le marqueur Y_{ijk} par une fonction de lien souple et paramétrée avant de le relier au processus latent :

$$H_k(Y_{ijk}; \boldsymbol{\eta}_k) = \tilde{Y}_{ijk} = \Lambda_{id}(t_{ij}) + \epsilon_{ijk}, \quad \forall k \in \mathcal{K}_d \ \& \ \forall j \in \tau_i \quad (3.6)$$

où H_k est une fonction de lien continue, dérivable et croissante paramétrée par le vecteur $\boldsymbol{\eta}_k$, \tilde{Y}_{ijk} est la version normalisée du marqueur Y_{ijk} et ϵ_{ijk} sont des erreurs de mesure indépendantes de variance σ_k^2 .

La fonction de lien monotone croissante peut être modélisée de manière souple à partir d'une base de splines intégrées (notés I-splines) [Ramsay, 1988]. Une I-spline est obtenue en intégrant une M-spline qui est une fonction non-négative de spline. Une combinaison de I-splines avec des coefficients positifs produit ainsi une fonction monotone croissante. Les I-splines avaient été utilisés précédemment pour modéliser de façon souple des fonctions de risque cumulé dans les analyses de survie [Joly, 1996].

En utilisant une base de I-splines de degré l avec p_k nœuds, $(\mathbb{I}_m)_{m=1, p_k+l}$, et des coefficients de régression associés positifs, on définit une fonction de lien H_k bijective croissante qui transforme le marqueur Y_{ijk} en un marqueur intermédiaire Gaussien \tilde{Y}_{ijk} . La fonction de lien H_k s'écrit :

$$H_k(Y_{ijk}; \boldsymbol{\eta}_k) = \tilde{Y}_{ijk} = \eta_{0k} + \sum_{m=1}^{p_k+l} \exp(\eta_{mk}) \mathbb{I}_m(Y_{ijk}),$$

avec $\boldsymbol{\eta}_k = (\eta_{mk})_{m=0, p_k+l}$ le vecteur de paramètres de transformation et $\exp(\eta_{mk})$ l'exponentielle du paramètre η_{mk} pour avoir des coefficients de régression associés à la base de I-splines positifs.

La Figure 3.1 donne un exemple de base de I-splines quadratiques et la fonction de lien correspondante pour un vecteur de paramètre donné.

3.1.2.3 Écriture matricielle du modèle d'observation

Il est possible d'écrire un modèle complet pour tous les marqueurs. Le vecteur de marqueurs transformés $\tilde{\mathbf{Y}}_{ij} = (\tilde{Y}_{ijk})_{k=1, K}^\top$ est relié au système de processus latents $\boldsymbol{\Lambda}_i(t_{ij})$ à l'aide de la matrice \mathbf{P} de taille $K \times D$ dont l'élément (k, d) vaut 1 si le marqueur k mesure le processus d :

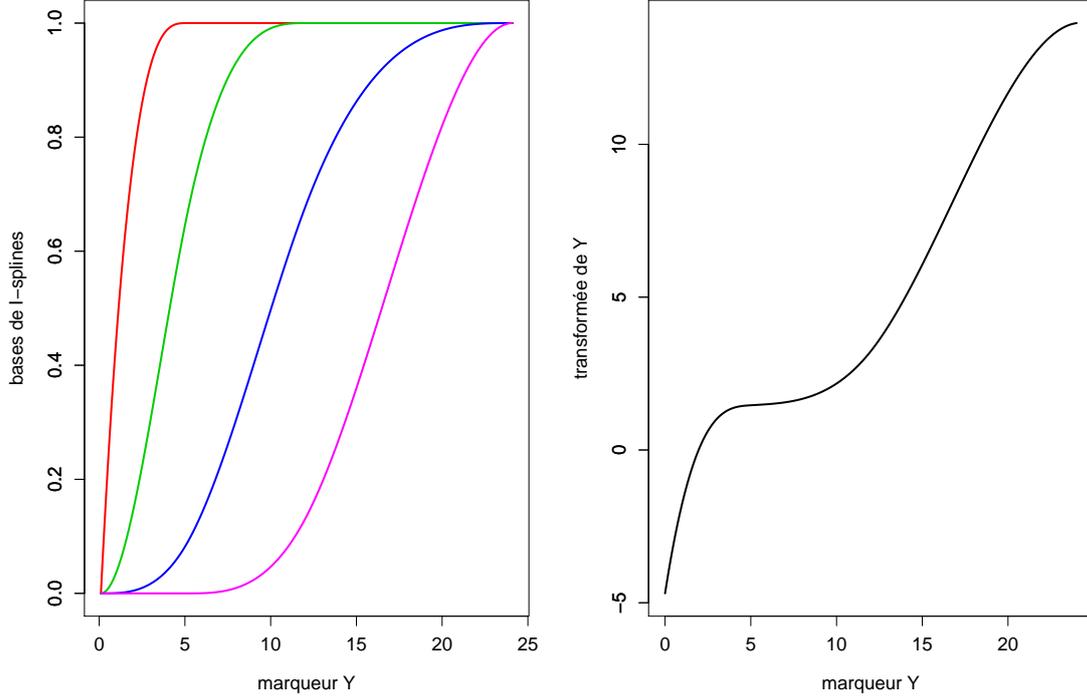


FIGURE 3.1 – Exemple d’une base de I-splines et d’une fonction de lien. La figure de gauche représente une base de splines intégrées de degré 2 avec 3 noeuds aux points 0, 5 et 12 utilisée pour transformer un marqueur à valeurs dans $[0, 24]$. La figure de droite représente la fonction de lien associée avec les paramètres de régression $(-4.70, \exp(1.80), \exp(-1.67), \exp(-2.06), \exp(2.48), \exp(-1.18))$.

$$\tilde{\mathbf{Y}}_{ij} = \mathbf{P}\boldsymbol{\Lambda}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}, \quad (3.7)$$

avec $\boldsymbol{\epsilon}_{ij} = (\epsilon_{ijk})_{k=1, \dots, K}^\top$ le vecteur des erreurs de mesure de variance-covariance $\boldsymbol{\Sigma} = \text{diag}((\sigma_k^2)_{k=1, \dots, K})$.

Dans la suite de la thèse, nous posons $\boldsymbol{\eta} = (\boldsymbol{\eta}_k^\top)_{k=1, \dots, K}^\top$, les paramètres utilisés pour transformer les K marqueurs, $\boldsymbol{\sigma} = (\sigma_k)_{k=1, \dots, K}$, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\sigma})$ le vecteur de tous les paramètres des modèles structurel et d’observation. Enfin nous notons $\phi_i(\tilde{\mathbf{Y}}_{ij} \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta})$ ou plus simplement $\phi_i(\tilde{\mathbf{Y}}_{ij})$, la densité multivariée du vecteur $\tilde{\mathbf{Y}}_{ij}$ conditionnellement aux variables explicatives $\mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}$ et au vecteur de paramètres $\boldsymbol{\theta}$.

3.1.3 Modèle d'observation pour données incomplètes

Dans la pratique, le processus d'observation peut être incomplet avec des données manquantes intermittentes pour un ou plusieurs marqueurs. En particulier, pour avoir un pas de discrétisation le plus fin possible, on peut ne pas choisir directement le pas de visite théorique des données mais un pas plus petit ce qui induit de fait des données manquantes intermittentes selon la définition du modèle d'observation donnée plus haut.

Cela se traduit au temps d'observation t_{ij} ($j \in \tau_i$) par un vecteur complet $\mathbf{Y}_{ij} = (\mathbf{Y}_{ij}^o, \mathbf{Y}_{ij}^m)$ avec \mathbf{Y}_{ij}^o le vecteur de taille K_{ij}^o des marqueurs observés au temps t_{ij} chez l'individu i et \mathbf{Y}_{ij}^m le vecteur de taille K_{ij}^m des marqueurs manquants ($K = K_{ij}^o + K_{ij}^m$). Notons $\mathbf{Y}_i^o = (\mathbf{Y}_{ij}^{o\top})_{j \in \tau_i}^\top$ le vecteur de taille K_i^o ($K_i^o = \sum_{j \in \tau_i} K_{ij}^o$) des mesures répétées incomplètes de l'individu i .

De manière analogue, le vecteur des observations transformées de l'individu i au temps t_{ij} s'écrit : $\tilde{\mathbf{Y}}_{ij} = (\tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m)$ et le vecteur de taille K_i^o des mesures répétées incomplètes transformées de l'individu i s'écrit : $\tilde{\mathbf{Y}}_i^o = (\tilde{\mathbf{Y}}_{ij}^{o\top})_{j \in \tau_i}^\top$.

Les données incomplètes intermittentes peuvent être reliées au système de processus latents à l'aide d'une matrice d'observation \mathbf{M}_{ij} de taille $K_{ij}^o \times K$ définie à chaque temps d'observation t_{ij} . L'élément (k^o, k) de la matrice \mathbf{M}_{ij} vaut 1 si le marqueur k est le $(k^o)^{\text{ième}}$ marqueur observé au temps t_{ij} et (k^o, k) vaut 0 sinon ; avec $k^o = 1, \dots, K_{ij}^o$ et $k = 1, \dots, K$:

$$\tilde{\mathbf{Y}}_{ij}^o = \mathbf{M}_{ij} \tilde{\mathbf{Y}}_{ij} = \mathbf{M}_{ij} \mathbf{P} \boldsymbol{\Lambda}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}^o \quad (3.8)$$

avec $\boldsymbol{\epsilon}_{ij}^o$ le vecteur d'erreurs de mesure indépendantes de matrice de variance $\mathbf{M}_{ij} \boldsymbol{\Sigma} \mathbf{M}_{ij}^\top$.

3.2 Estimation du modèle

Les paramètres du modèle contenus dans le vecteur $\boldsymbol{\theta}$ sont estimés simultanément par la méthode de maximum de vraisemblance. En vue d'établir l'expression de la vraisemblance du modèle, nous avons commencé par établir les expressions des distributions marginales du système de processus latents et des observations transformées.

3.2.1 Distributions des processus latents et des observations transformées

À partir de l'expression du taux de variation du système de processus latents, on peut déduire l'expression de la valeur du système au temps $t_{ij} + \delta$ en fonction de sa valeur au temps t_{ij} :

$$\begin{aligned}\boldsymbol{\Lambda}_i(t_{ij} + \delta) &= \delta(\mathbf{X}_i(t_{ij} + \delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t_{ij} + \delta)\mathbf{v}_i) + (\delta\mathbf{A}_{i,\delta}(t_{ij}) + \mathbb{I}_D)\boldsymbol{\Lambda}_i(t_{ij}), \quad \forall t_{ij} \geq 0 \\ &= \delta(\mathbf{X}_i(t_{ij} + \delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t_{ij} + \delta)\mathbf{v}_i) + \tilde{\mathbf{A}}_{i,\delta}(t_{ij})\boldsymbol{\Lambda}_i(t_{ij}),\end{aligned}\quad (3.9)$$

avec $\tilde{\mathbf{A}}_{i,\delta}(t_{ij}) = \delta\mathbf{A}_{i,\delta}(t_{ij}) + \mathbb{I}_D$; \mathbb{I}_D la matrice identité de taille D .

En posant $t_{ij} = j\delta$, $t_{ij} + \delta = t_{i(j+1)}$, on a que :

$$\boldsymbol{\Lambda}_i(t_{ij}) = \delta(\mathbf{X}_i(t_{ij})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{ij})\mathbf{v}_i) + \tilde{\mathbf{A}}_{i,\delta}(t_{i(j-1)})\boldsymbol{\Lambda}_i(t_{i(j-1)}), \quad \forall j \in \{1, \dots, J\}.$$

Pour $j = 1$, $\boldsymbol{\Lambda}_i(t_{i1}) = \delta(\mathbf{X}_i(t_{i1})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{i1})\mathbf{v}_i) + \tilde{\mathbf{A}}_{i,\delta}(0)\boldsymbol{\Lambda}_i(0)$.

Pour $j = 2$, $\boldsymbol{\Lambda}_i(t_{i2}) = \delta(\mathbf{X}_i(t_{i2})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{i2})\mathbf{v}_i) + \tilde{\mathbf{A}}_{i,\delta}(t_{i1})(\mathbf{X}_i(t_{i1})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{i1})\mathbf{v}_i) + \tilde{\mathbf{A}}_{i,\delta}(t_{i1})\tilde{\mathbf{A}}_{i,\delta}(0)\boldsymbol{\Lambda}_i(0)$

Par récurrence, l'expression du système de processus latents au temps d'observation t_{ij} , $\boldsymbol{\Lambda}_i(t_{ij})$ s'écrit :

$$\left\{ \begin{array}{ll} \boldsymbol{\Lambda}_i(t_{ij}) = \mathbf{X}_i^{(0)}\boldsymbol{\beta} + \mathbf{u}_i & \text{si } j = 0 \\ \boldsymbol{\Lambda}_i(t_{ij}) = \tilde{\mathbf{A}}_{i,\delta}(0)(\mathbf{X}_i^{(0)}\boldsymbol{\beta} + \mathbf{u}_i) & \text{si } j = 1 \\ \boldsymbol{\Lambda}_i(t_{ij}) = \prod_{l=0}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_{il})(\mathbf{X}_i^{(0)}\boldsymbol{\beta} + \mathbf{u}_i) & \text{si } j > 1 \\ \quad + \delta(\mathbf{X}_i(t_{ij})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{ij})\mathbf{v}_i) & \\ \quad + \delta \sum_{s=1}^{j-1} \prod_{l=s}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_{il})(\mathbf{X}_i(t_{is})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{is})\mathbf{v}_i) & \end{array} \right. \quad (3.10)$$

où $t_{ij} = j \times \delta$ pour $j \in \{0, \dots, J\}$.

En introduisant $\boldsymbol{\Psi}_{i,\delta}(s, j)$ pour $s \leq j$ tels que

$$\boldsymbol{\Psi}_{i,\delta}(s, j) = \begin{cases} \mathbf{I}_D, & \text{si } s = j \\ \prod_{l=s}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_{il}) & \text{si } s < j \end{cases}, \quad (3.11)$$

MÉTHODOLOGIE

L'équation (3.10) devient :

$$\Lambda_i(t_{ij}) = \Psi_{i,\delta}(0, j)(\mathbf{X}_i^{(0)}\beta + \mathbf{u}_i) + \left(\delta \sum_{s=1}^j \Psi_{i,\delta}(s, j) (\mathbf{X}_i(t_{is})\gamma + \mathbf{Z}_i(t_{is})\mathbf{v}_i) \right) \mathbb{1}_{\{j>0\}} \quad (3.12)$$

Sous l'hypothèse de normalité du vecteur des effets aléatoires \mathbf{w}_i , le vecteur de processus latents $\Lambda_i(t_{ij})$ au temps discret t_{ij} et le vecteur $\Lambda_i = \left(\Lambda_i(t_{ij})^\top \right)_{j \in \tau_i}^\top$ de tous les états du système de processus latents sont Gaussiens d'espérances respectives $\boldsymbol{\mu}_{\Lambda_{ij}}$ et $\boldsymbol{\mu}_{\Lambda_i} = \left(\boldsymbol{\mu}_{\Lambda_{ij}}^\top \right)_{j \in \tau_i}^\top$ et de matrices de variance-covariance respectives $\mathbf{V}_{\Lambda_{ijj}} = \text{var}(\Lambda_i(t_{ij}))$ et $\mathbf{V}_{\Lambda_i} = \left(\mathbf{V}_{\Lambda_{ijj'}} \right)_{(j,j') \in \tau_i^2}$ avec

$$\boldsymbol{\mu}_{\Lambda_{ij}} = \mathbb{E}[\Lambda_i(t_{ij})] = \Psi_{i,\delta}(0, j)\mathbf{X}_i^{(0)}\beta + \left(\delta \sum_{s=1}^j \Psi_{i,\delta}(s, j)\mathbf{X}_i(t_{is})\gamma \right) \mathbb{1}_{\{j>0\}} \quad (3.13)$$

et

$$\begin{aligned} \mathbf{V}_{\Lambda_{ijj'}} &= \text{cov}(\Lambda_i(t_{ij}); \Lambda_i(t_{ij'})) \\ &= \Psi_{i,\delta}(0, j)\mathbf{B}_u\Psi_{i,\delta}(0, j')^\top \\ &\quad + \left(\Psi_{i,\delta}(0, j)\mathbf{B}_{uv} \left(\delta \sum_{s'=1}^{j'} \Psi_{i,\delta}(s', j')\mathbf{Z}_i(t_{is'}) \right)^\top \right) \mathbb{1}_{\{j'>0\}} \\ &\quad + \left(\left(\delta \sum_{s=1}^j \Psi_{i,\delta}(s, j)\mathbf{Z}_i(t_{is}) \right) \mathbf{B}_{uv}^\top \Psi_{i,\delta}(0, j')^\top \right) \mathbb{1}_{\{j>0\}} \\ &\quad + \left(\left(\delta \sum_{s=1}^j \Psi_{i,\delta}(s, j)\mathbf{Z}_i(t_{is}) \right) \mathbf{B}_v \left(\delta \sum_{s'=1}^{j'} \Psi_{i,\delta}(s', j')\mathbf{Z}_i(t_{is'}) \right)^\top \right) \mathbb{1}_{\{\min(j,j')>0\}} \end{aligned} \quad (3.14)$$

Distribution des données incomplètes

Le vecteur des données incomplètes transformées $\tilde{\mathbf{Y}}_{ij}^o$ au temps d'observation t_{ij} est Gaussien d'espérance $\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_{ij}^o} = \mathbf{M}_{ij}\mathbf{P}\boldsymbol{\mu}_{\Lambda_{ij}}$ et de matrice de variance-covariance $\mathbf{V}_{\tilde{\mathbf{Y}}_{ij}^o} = \mathbf{M}_{ij}(\mathbf{P}\mathbf{V}_{\Lambda_{ijj}}\mathbf{P}^\top + \boldsymbol{\Sigma})\mathbf{M}_{ij}^\top$; et le vecteur des données incomplètes de l'individu i $\tilde{\mathbf{Y}}_i^o = \left(\tilde{\mathbf{Y}}_{ij}^{o\top} \right)_{j \in \tau_i}^\top$ suit une distribution Gaussienne d'espérance $\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o} = \left(\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_{ij}^o}^\top \right)_{j \in \tau_i}^\top$ et de matrice de variance-covariance $\mathbf{V}_{\tilde{\mathbf{Y}}_i^o}$, une matrice de $(n_i + 1) \times (n_i + 1)$ blocs dont le bloc (j, j') vaut $\mathbf{M}_{ij}\mathbf{P}\mathbf{V}_{\Lambda_{ijj'}}\mathbf{P}^\top\mathbf{M}_{ij'}^\top + \left(\mathbf{M}_{ij}\boldsymbol{\Sigma}\mathbf{M}_{ij'}^\top \right) \mathbb{1}_{\{j=j'\}}$.

3.2.2 Identifiabilité du modèle

Nous revenons dans cette section sur les conditions d'identifiabilité du modèle.

3.2.2.1 Modèle identifiable : définition

Soit un modèle $M : (Y_i \in \Omega, \mathbb{P}_\theta, \theta \in \Theta)$, où :

- Ω l'espace de réalisation des variables aléatoires Y_i
- Θ l'espace des valeurs possibles pour le paramètre θ
- \mathbb{P}_θ une loi de probabilité de densité f_θ .

Le modèle M est identifiable si et seulement si l'application qui à θ associe f_θ est injective, c'est-à-dire si et seulement si : $\forall \theta \neq \theta' \Rightarrow f_\theta(Y_i) \neq f_{\theta'}(Y_i)$.

3.2.2.2 Evaluation de l'identifiabilité du modèle

D'après l'expression de la distribution des processus latents établie en section 3.2.1 chaque processus latent $\mathbf{\Lambda}_{id} = (\Lambda_{id}(0), \Lambda_{id}(t_{i1}), \dots, \Lambda_{id}(t_{iJ}))$ est un vecteur Gaussien $\mathbf{\Lambda}_{id} \sim \mathcal{N}(\boldsymbol{\mu}_{id}, \mathbf{V}_{id})$. Pour un marqueur Gaussien Y_{ijk} relié au processus latent $\mathbf{\Lambda}_{id}$ par le modèle d'observation 3.4 de la section 3.1.2, le modèle complet du processus $\mathbf{\Lambda}_{id}$ et du marqueur Y_{ijk} s'écrit :

$$\begin{cases} \mathbf{\Lambda}_{id} \sim \mathcal{N}(\boldsymbol{\mu}_{id}, \mathbf{V}_{id}) \\ Y_{ijk} \sim \mathcal{N}(\eta_{0k} + \eta_{1k}\mu_{ijd} ; \eta_{1k}^2 V_{ijd} + \sigma_k^2), j = 0, \dots, J; \forall k \in \mathcal{K}_d \end{cases} \quad (3.15)$$

avec μ_{ijd} et V_{ijd} l'espérance et la variance du processus latent au temps t_{ij} , σ_k^2 la variance des erreurs de mesure Gaussiennes indépendantes ϵ_{ijk} .

A partir du moment où le modèle structurel n'a pas été contraint, l'expression 3.15 peut être réécrite pour faire ressortir l'intercept du modèle du processus Λ_d noté β_{0d} qui intervient dans la définition du niveau initial du système, et la variance de l'intercept aléatoire u_{id} notée ici σ_d^2 qui intervient aussi dans la définition du niveau initial du processus :

$$\begin{cases} \mathbf{\Lambda}_{id} \sim \mathcal{N}(\beta_{0d}\mathbf{1}_{n_i+1} + \boldsymbol{\mu}_{id}^0, \sigma_d^2 \mathbf{V}_{id}^0) \\ Y_{ijk} \sim \mathcal{N}(\eta_{0k} + \eta_{1k}(\beta_{0d} + \mu_{ijd}^0) ; \eta_{1k}^2(\sigma_d^2 V_{ijd}^0) + \sigma_k^2), j = 0, \dots, J; \forall k \in \mathcal{K}_d \end{cases} \quad (3.16)$$

avec μ_{id}^0 la moyenne de Λ_{id} privée de son intercept, V_{id}^0 la variance de Λ_{id} standardisée par rapport à la variance de son intercept aléatoire, $\mathbf{1}_n$ le vecteur de taille n contenant des 1.

On peut alors voir aisément que pour tout couple $(\beta_{0d}^*, \sigma_d^*) \neq (\beta_{0d}, \sigma_d)$, on trouvera \mathcal{K}_d couples $(\eta_{0k}^*, \eta_{1k}^*)$ ($k = 1, \dots, \mathcal{K}_d$) tels que la distribution de Y_{ijk} reste identique :

$$\begin{cases} E_{\boldsymbol{\theta}}[Y_{ijk}] = E_{\boldsymbol{\theta}^*}[Y_{ijk}] \\ \text{Var}_{\boldsymbol{\theta}}[Y_{ijk}] = \text{Var}_{\boldsymbol{\theta}^*}[Y_{ijk}] \end{cases} \quad (3.17)$$

MÉTHODOLOGIE

où θ et θ^* ne diffèrent que par les valeurs prises par (β_{0d}, σ_d) et (η_{0k}, η_{1k}) pour $k = 1, \dots, \mathcal{K}_d$

Les densités des observations Y_{ijk} pour les vecteurs de paramètres θ et θ^* sont donc égales : $\theta \neq \theta^*$ et $f_\theta(Y_{ijk}) = f_{\theta^*}(Y_{ijk})$. De même, les densités du vecteur d'observations $\mathbf{Y}_{ik} = (Y_{i01}, \dots, Y_{iJk})$ pour θ et θ^* sont égales : $\theta \neq \theta^*$, $f_\theta(Y_{ik}) = f_{\theta^*}(Y_{ik})$. Le modèle 3.16 n'est donc pas identifiable.

Ce problème d'identifiabilité est bien connu dans l'approche de modélisation à variable latente et il n'arrive pas uniquement dans le cas d'équations d'observations linéaires. Lorsque le modèle d'observation n'est pas contraint ou que la dimension de la variable latente n'est pas fixée, le modèle est globalement non identifiable [Sánchez et al., 2005].

Pour rendre notre modèle identifiable, nous avons fixé la dimension de chaque processus Λ_{id} en rajoutant une contrainte sur la localisation μ_{id} et une contrainte sur la dispersion V_{id} . Nous avons contraint l'intercept et la variance du niveau initial des processus latents respectivement à 0 et 1 c'est-à-dire l'intercept du niveau initial $\beta_{0d} = 0$ et la variance de l'intercept aléatoire sur le niveau initial $\text{Var}(u_i) = 1$. De cette façon là, les modèles d'observations restent non contraints.

Dans la suite, le vecteur total des paramètres θ à estimer exclut donc les intercepts des processus et la variance des intercepts aléatoires u_d .

3.2.3 Log-vraisemblance du modèle

Pour un échantillon de N sujets indépendants et identiquement distribués, la log-vraisemblance $L(\mathbf{Y}^o, \theta)$ du modèle est la somme des contributions individuelles $L_i(\mathbf{Y}^o, \theta)$:

$$L(\mathbf{Y}^o, \theta) = \sum_{i=1}^N L_i(\mathbf{Y}_i^o, \theta) = \sum_{i=1}^N \log(\mathcal{L}_i(\mathbf{Y}^o, \theta)) \quad (3.18)$$

où $\mathcal{L}_i(\mathbf{Y}^o, \theta)$ est la contribution de l'individu i à la vraisemblance du modèle :

$$\mathcal{L}_i(\mathbf{Y}_i^o; \theta) = \phi_i(\tilde{\mathbf{Y}}_i^o; \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o}, \mathbf{V}_{\tilde{\mathbf{Y}}_i^o}) \prod_{j \in \tau_i} \prod_{l=1}^{K_{ij}^o} \mathcal{J}_{k(l)} \left(H_{k(l)}(\tilde{\mathbf{Y}}_{ijk(l)}^o; \boldsymbol{\eta}_{k(l)}) \right) \quad (3.19)$$

où

$$\phi_i(\tilde{\mathbf{Y}}_i^o; \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o}, \mathbf{V}_{\tilde{\mathbf{Y}}_i^o}) = (2\pi)^{-\frac{K_i^o}{2}} |\mathbf{V}_{\tilde{\mathbf{Y}}_i^o}|^{-\frac{1}{2}} e^{-\frac{1}{2} (\tilde{\mathbf{Y}}_i^o - \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o})^\top (\mathbf{V}_{\tilde{\mathbf{Y}}_i^o})^{-1} (\tilde{\mathbf{Y}}_i^o - \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o})}$$

est la densité multivarié Gaussienne d'espérance $\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o}$ et de matrice de variance-covariance $\mathbf{V}_{\tilde{\mathbf{Y}}_i^o}$ et $\mathcal{J}_{k(l)}(H_{k(l)}(\tilde{\mathbf{Y}}_{ijk(l)}^o; \boldsymbol{\eta}_{k(l)}))$ est le Jacobien de la fonction de lien $H_{k(l)}$ utilisé pour

transformer le marqueur $k(l)$ correspondant au $l^{\text{ième}}$ marqueur observé au temps t_{ij} chez l'individu i . Par exemple, pour une fonction de lien linéaire, $\mathcal{J}_{k(l)} \left(H_{k(l)}(\tilde{Y}_{ijk(l)}^o; \eta_{k(l)}) \right) = \frac{1}{\eta_{1k(l)}}$. Notons que la log-vraisemblance de notre modèle est une fonction non-linéaire du vecteur de paramètres $\boldsymbol{\theta}$.

Hypothèse de données manquantes aléatoires

L'estimateur du maximum de vraisemblance du vecteur de paramètres $\boldsymbol{\theta}$ obtenu à partir des données incomplètes \mathbf{Y}^o est un estimateur consistant du vecteur de paramètres $\boldsymbol{\theta}$ du modèle complet sous l'hypothèse que les données sont manquantes aléatoirement (MAR pour « missing at random »).

En effet, notons $(R_{ij})_{j \in \tau_i}$ le processus d'observation tel que, au temps t_{ij} , $R_{ij} = 1$ si tous les marqueurs sont observés (i.e $K_{ij}^m = 0$) et $R_{ij} = 0$ si au moins un marqueur est manquant. Nous supposons que la loi de probabilité de R_{ij} est caractérisée par un vecteur de paramètres $\boldsymbol{\xi}$ distinct de $\boldsymbol{\theta}$.

La densité des données incomplètes (observées) sachant les variables explicatives et les paramètres des modèles est :

$$\begin{aligned} f(\tilde{\mathbf{Y}}_{ij}^o, R_{ij} \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) &= \int_{\mathbb{R}} f(\tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m, R_{ij} \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\xi}) d\tilde{\mathbf{Y}}_{ij}^m \\ &= \int_{\mathbb{R}} f(R_{ij} \mid \tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m, \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\xi}) f(\tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) d\tilde{\mathbf{Y}}_{ij}^m \end{aligned} \quad (3.20)$$

Sous l'hypothèse de données manquantes aléatoirement, $R_{ij} \perp\!\!\!\perp \tilde{\mathbf{Y}}_{ij}^m \mid \tilde{\mathbf{Y}}_{ij}^o$, c'est-à-dire :

$$f(R_{ij} \mid \tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m, \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\xi}) = f(R_{ij} \mid \tilde{\mathbf{Y}}_{ij}^o, \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\xi})$$

L'expression 3.20 devient :

$$f(\tilde{\mathbf{Y}}_{ij}^o, R_{ij} \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) = f(R_{ij} \mid \tilde{\mathbf{Y}}_{ij}^o, \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\xi}) \int_{\mathbb{R}} f(\tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) d\tilde{\mathbf{Y}}_{ij}^m,$$

Et la vraisemblance des données observées $\tilde{\mathbf{Y}}^o$ basée sur les densités

$$f(\tilde{\mathbf{Y}}_{ij}^o \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) = \int_{\mathbb{R}} f(\tilde{\mathbf{Y}}_{ij}^o, \tilde{\mathbf{Y}}_{ij}^m \mid \mathbf{X}^{(0)}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}) d\tilde{\mathbf{Y}}_{ij}^m \quad (3.21)$$

pour $i = 1, \dots, N$ et $j = 1, \dots, n_i$ permet donc d'obtenir l'estimateur du maximum de vraisemblance de $\boldsymbol{\theta}$.

3.2.4 Maximisation de la log-vraisemblance

3.2.4.1 Écriture du problème d'optimisation

L'estimateur du maximum de vraisemblance du vecteur de paramètres $\boldsymbol{\theta}$ est obtenu en maximisant la log-vraisemblance $L(\mathbf{Y}^o; \boldsymbol{\theta})$ du modèle. Le problème d'optimisation s'écrit :

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} L(\mathbf{Y}^o; \boldsymbol{\theta}) \quad (3.22)$$

Le problème d'optimisation n'admet pas de solution analytique. Nous avons recours à une résolution itérative qui consiste à partir d'un point initial $\boldsymbol{\theta}^{(0)}$ puis à progressivement améliorer la solution $\boldsymbol{\theta}^{(k)}$ à l'itération k : $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \Delta\boldsymbol{\theta}^{(k)}$, où $\Delta\boldsymbol{\theta}^{(k)}$ est l'incrément à l'itération k , jusqu'à converger au bout d'un certain nombre d'itérations vers la solution numérique.

3.2.4.2 Algorithme d'optimisation

Nous avons utilisé une extension de l'algorithme de Levenberg-Marquardt [Marquardt, 1963] qui est adapté pour l'optimisation de fonctions non-linéaires. L'algorithme de Levenberg-Marquardt est une extension de la méthode d'optimisation de Gauss-Newton. La méthode de Gauss-Newton consiste à linéariser la fonction à optimiser (c'est-à-dire la log-vraisemblance) grâce au développement en série de Taylor et ensuite à maximiser la version linéarisée de la fonction à optimiser, dans notre cas, $L^*(\mathbf{Y}^o; \boldsymbol{\theta})$, la version linéarisée de la log-vraisemblance $L(\mathbf{Y}^o; \boldsymbol{\theta})$, soit :

$$L(\mathbf{Y}^o; \boldsymbol{\theta}) \approx L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) + \nabla L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) = L^*(\mathbf{Y}^o; \boldsymbol{\theta}) \quad (3.23)$$

où $\nabla L(\mathbf{Y}^o; \boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta}) = \left(\frac{\partial L(\mathbf{Y}^o; \boldsymbol{\theta})}{\partial \theta_1}, \dots, \frac{\partial L(\mathbf{Y}^o; \boldsymbol{\theta})}{\partial \theta_p} \right)^\top$, avec p le nombre de paramètres du vecteur $\boldsymbol{\theta}$.

La version linéarisée $L^*(\mathbf{Y}^o; \boldsymbol{\theta})$ de la log-vraisemblance est maximisée en résolvant l'équation :

$$\begin{aligned} \nabla L^*(\mathbf{Y}^o; \boldsymbol{\theta}) &= 0 \\ \Leftrightarrow \nabla^2 L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)}) + \nabla L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) &= 0 \\ \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)} - \left(\nabla^2 L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) \right)^{-1} \nabla L(\mathbf{Y}^o; \boldsymbol{\theta}^{(k)}) \\ \Leftrightarrow \boldsymbol{\theta} = \boldsymbol{\theta}^{(k)} - \left(\mathbf{H}(\boldsymbol{\theta}^{(k)}) \right)^{-1} \mathbf{U}(\boldsymbol{\theta}^{(k)}) \end{aligned} \quad (3.24)$$

où $\mathbf{H}(\boldsymbol{\theta}^{(k)})$ et $\mathbf{U}(\boldsymbol{\theta}^{(k)})$ sont respectivement la matrice Hessienne et le gradient de la log-vraisemblance à l'itération k . La solution du problème d'optimisation est construite

de manière itérative comme suit :

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \left(\mathbf{H}(\boldsymbol{\theta}^{(k)}) \right)^{-1} \mathbf{U}(\boldsymbol{\theta}^{(k)}) \quad (3.25)$$

Dans la méthode de Gauss-Newton, la matrice Hessienne n'est pas toujours inversible et au voisinage de $\hat{\boldsymbol{\theta}}$, l'algorithme converge lentement. L'algorithme de Levenberg-Marquardt [Marquardt, 1963] améliore la méthode de Gauss-Newton en gonflant au besoin la diagonale de la matrice Hessienne à chaque itération k à l'aide du paramètre $\lambda^{(k)}$ et en considérant une fraction de l'incrément $\Delta\boldsymbol{\theta}^{(k)}$ grâce à un paramètre ν :

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - \nu \left(\mathbf{H}(\boldsymbol{\theta}^{(k)}) + \lambda^{(k)} \mathbf{I} \right)^{-1} \mathbf{U}(\boldsymbol{\theta}^{(k)}). \quad (3.26)$$

Les critères de convergence de l'algorithme de Levenberg-Marquardt sont relatifs à :

- la stabilité de la log-vraisemblance : $|L(\mathbf{Y}^*; \boldsymbol{\theta}^{(k+1)}) - L(\mathbf{Y}^*; \boldsymbol{\theta}^{(k)})| < \epsilon_a$,
- la stabilité du vecteur de paramètres : $\|\boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}\|_2 < \epsilon_b$,
- la stabilité de l'erreur numérique : $\frac{\mathbf{U}(\boldsymbol{\theta}^{(k)})^\top [\mathbf{H}^{(p)}]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(k)})}{p} < \epsilon_c$

avec p le nombre total de paramètres et ϵ_a , ϵ_b , ϵ_c des seuils fixés par l'utilisateur que nous avons tous fixés à $1e - 4$. Le troisième critère $\frac{\mathbf{U}(\boldsymbol{\theta}^{(k)})^\top [\mathbf{H}^{(p)}]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(k)})}{p} < \epsilon_c$ définit la stabilité de l'erreur numérique.

3.3 Relations entre matrices d'influences temporelles sous différents pas de discrétisation

Dans notre approche de modélisation, les processus sont définis en temps discret avec un pas de discrétisation δ . Ce choix de modélisation en temps discret est fait pour éviter les problèmes d'estimation rencontrés dans les modèles d'équations différentielles évoqués dans l'état de l'art (section 2.2.4.6). Dans les modèles d'équations différentielles, ce sont les influences temporelles au niveau infinitésimal qui sont modélisées entre les processus latents bien que les estimations soient réalisées à partir d'observations en temps discrets.

Pour mieux comprendre l'interprétation des matrices d'influences temporelles dans notre approche, nous avons évalué l'impact de la discrétisation du temps sur la structure infinitésimale d'influences temporelles entre les processus latents par des simulations décrites dans le chapitre 4. Mais nous avons aussi exploré cet impact de façon analytique en établissant la relation entre la matrice d'influences temporelles en temps discret et la matrice d'influences temporelles au niveau infinitésimal (i.e en temps continu).

Relation entre les structures d'influences temporelles pour deux pas différents de discrétisation

Nous avons d'abord commencé par établir la relation entre les matrices de transition obtenues pour deux pas de discrétisation : δ^o et $\delta = \frac{\delta^o}{\rho}$ avec ρ un nombre entier supérieur à l'unité ($\rho > 1$).

Reprenons l'expression de la valeur du système de processus latents au temps $t + \delta$

$$\mathbf{\Lambda}_i(t + \delta) = \delta(\mathbf{X}_i(t + \delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta)\mathbf{v}_i) + \tilde{\mathbf{A}}_{i\delta}(t)\mathbf{\Lambda}_i(t).$$

où $\tilde{\mathbf{A}}_{i,\delta}(t) = \delta\mathbf{A}_{i,\delta}(t) + \mathbb{I}_D$.

Avec un pas de discrétisation $\delta = \frac{\delta^o}{\rho}$ (avec $\rho > 1$), le système de processus latents évolue du temps t jusqu'au temps $t + \rho\delta$ où l'expression du système établie par récurrence vaut :

$$\begin{aligned} \mathbf{\Lambda}_i(t + \rho\delta) &= \delta(\delta\mathbf{A}_{i,\delta}(t + \rho\delta) + \mathbb{I}_D) (\mathbf{X}_i(t + \rho\delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \rho\delta)\mathbf{v}_i) \\ &\quad + \delta \sum_{s=1}^{\rho-1} \prod_{l=s}^{\rho-1} (\delta\mathbf{A}_{i,\delta}(t + \delta l) + \mathbb{I}_D) (\mathbf{X}_i(t + \delta s)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta s)\mathbf{v}_i) \\ &\quad + \prod_{l=0}^{\rho-1} (\delta\mathbf{A}_{i,\delta}(t + \delta l) + \mathbb{I}_D)\mathbf{\Lambda}_i(t). \end{aligned} \tag{3.27}$$

On peut aussi considérer directement le pas de discrétisation δ^o , auquel cas la seconde ligne de l'équation (3.1) peut se réécrire :

$$\mathbf{\Lambda}_i(t + \delta^o) = \delta^o (\mathbf{X}_i(t + \delta^o)\boldsymbol{\gamma}^o + \mathbf{Z}_i(t + \delta^o)\mathbf{v}_i^o) + (\mathbb{I}_D + \delta^o\mathbf{A}_{i,\delta^o}(t)) \mathbf{\Lambda}_i(t). \tag{3.28}$$

Pour relier les matrices d'influences temporelles sous les deux pas de discrétisation δ^o et $\delta = \frac{\delta^o}{\rho}$ nous faisons l'hypothèse raisonnable que sur l'intervalle de temps ouvert $]t, t + \delta^o[$, la variation du système de processus latents est sensiblement la même pour les deux pas de discrétisation i.e :

$$\begin{aligned}
 & \left[\begin{aligned}
 & \delta(\delta \mathbf{A}_{i,\delta}(t + \delta l) + \mathbb{I}_D) (\mathbf{X}_i(t + \delta \rho)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta \rho)\mathbf{v}_i) \\
 & + \delta \sum_{s=1}^{\rho-1} \prod_{l=s}^{\rho-1} (\delta \mathbf{A}_{i,\delta}(t + \delta l) + \mathbb{I}_D) (\mathbf{X}_i(t + \delta s)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta s)\mathbf{v}_i) \\
 & + \prod_{l=0}^{\rho-1} (\delta \mathbf{A}_{i,\delta}(t + \delta l) + \mathbb{I}_D) \boldsymbol{\Lambda}_i(t)
 \end{aligned} \right] \\
 & \approx \\
 & \delta^\circ (\mathbf{X}_i(t + \delta^\circ)\boldsymbol{\gamma}^\circ + \mathbf{Z}_i(t + \delta^\circ)\mathbf{v}_i^\circ)
 \end{aligned}$$

Sous cette hypothèse, on obtient en égalant les seconds membres des équations (3.27) et (3.28)

$$\begin{aligned}
 \mathbb{I}_D + \delta^\circ \mathbf{A}_{i,\delta^\circ}(t) & \approx \prod_{l=0}^{\rho-1} (\mathbb{I}_D + \delta \mathbf{A}_{i,\delta}(t + l\delta)) \\
 \mathbf{A}_{i,\delta^\circ}(t) & \approx \frac{1}{\delta^\circ} \left(\prod_{l=0}^{\rho-1} \left(\mathbb{I}_D + \frac{\delta^\circ}{\rho} \mathbf{A}_{i,\frac{\delta^\circ}{\rho}}(t + l\frac{\delta^\circ}{\rho}) \right) - \mathbb{I}_D \right) \quad (3.29)
 \end{aligned}$$

Supposons que la structure d'influences temporelles entre les processus latents $\mathbf{A}_{i,\frac{\delta^\circ}{\rho}}(t)$ est constante sur l'intervalle $[t, t + \delta^\circ]$, la relation(3.29) devient :

$$\mathbf{A}_{i,\delta^\circ}(t) \approx \frac{1}{\delta^\circ} \left(\left(\mathbb{I}_D + \frac{\delta^\circ}{\rho} \mathbf{A}_{i,\frac{\delta^\circ}{\rho}}(t) \right)^\rho - \mathbb{I}_D \right) \quad (3.30)$$

Relation entre la structure d'influences temporelles en temps discret et la structure d'influences temporelles au niveau infinitésimal

Supposons que pour δ° fixé et fini, la limite $\lim_{\rho \rightarrow \infty} \mathbf{A}_{i,\frac{\delta^\circ}{\rho}}(t)$ existe et est finie :

$$\lim_{\rho \rightarrow \infty} \mathbf{A}_{i,\frac{\delta^\circ}{\rho}}(t) = \mathbf{A}_i(t)$$

La limite à l'infini du second membre de l'égalité (3.30) donne la relation entre la structure de dépendance temporelle $\mathbf{A}_{i,\delta^\circ}(t)$ obtenue avec un pas de discrétisation δ° et la structure de dépendance temporelle $\mathbf{A}_i(t)$ en temps continu :

$$\begin{aligned}
\mathbf{A}_{i,\delta^o}(t) &\approx \lim_{\rho \rightarrow \infty} \frac{1}{\delta^o} \left(\left(\mathbf{I}_D + \frac{\delta^o}{\rho} \mathbf{A}_{i,\frac{\delta^o}{\rho}}(t) \right)^\rho - \mathbf{I}_D \right) \\
\mathbf{A}_{i,\delta^o}(t) &\approx \frac{1}{\delta^o} \left(\exp \left(\delta^o \lim_{\rho \rightarrow \infty} \mathbf{A}_i(t) \right) - \mathbf{I}_D \right) \\
\mathbf{A}_{i,\delta^o}(t) &\approx \frac{1}{\delta^o} \left(\exp(\delta^o \mathbf{A}_i(t)) - \mathbf{I}_D \right)
\end{aligned} \tag{3.31}$$

En particulier lorsque δ^o est très petit, on retrouve bien

$$\mathbf{A}_{i,\delta^o}(t) \approx \mathbf{A}_i(t) \tag{3.32}$$

3.4 Prédictions

Les prédictions marginales et les prédictions spécifiques aux sujets issues du modèle peuvent être obtenues dans l'échelle des processus latents et dans l'échelle naturelle des marqueurs.

3.4.1 Prédictions des processus latents

Prédictions marginales

Les prédictions marginales du système de processus latents $\hat{\Lambda}_i^{(M)}$ du sujet i s'obtiennent à partir de la distribution marginale du système de processus latents :

$$\hat{\Lambda}_i^{(M)} = \mathbb{E}[\Lambda_i] = \left(\left(\boldsymbol{\mu}_{ij} \right)_{j \in \tau_i}^\top \right)^\top \tag{3.33}$$

Prédictions spécifiques au sujet i

Les prédictions spécifiques au sujet i $\hat{\Lambda}_i^{(C)}$ sont obtenues à partir de la distribution du système de processus latents conditionnelle aux observations \mathbf{Y}_i^o de l'individu i ou de façon équivalente aux observations transformées $\tilde{\mathbf{Y}}_i^o$. Elles s'établissent à partir de la distribution jointe du vecteur $\left(\Lambda_i^\top, \tilde{\mathbf{Y}}_i^{o\top} \right)^\top$. En effet, le vecteur $\left(\Lambda_i^\top, \tilde{\mathbf{Y}}_i^{o\top} \right)^\top$ suit une loi normale :

$$\begin{pmatrix} \Lambda_i \\ \tilde{\mathbf{Y}}_i^o \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \boldsymbol{\mu}_{\Lambda_i} \\ \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^o} \end{pmatrix}, \begin{pmatrix} \mathbf{V}_{\Lambda_i} & \mathbf{C}_{\Lambda_i \tilde{\mathbf{Y}}_i^o} \\ \mathbf{C}_{\Lambda_i \tilde{\mathbf{Y}}_i^o}^\top & \mathbf{V}_{\tilde{\mathbf{Y}}_i^o} \end{pmatrix} \right) \tag{3.34}$$

avec $\mathbf{C}_{\Lambda_i \tilde{\mathbf{Y}}_i^o} = \left(\hat{\mathbf{V}}_{\Lambda_i j j'} \mathbf{P}^\top \mathbf{M}_{ij'}^\top \right)_{(j,j') \in \tau_i^2}$ la matrice de variance-covariance entre Λ_i et $\tilde{\mathbf{Y}}_i^o$.

À partir de l'équation (3.34) la distribution du vecteur Λ_i conditionnelle au vecteur \tilde{Y}_i^o est normale d'espérance $\mu_{\Lambda_i} + C_{\Lambda_i \tilde{Y}_i^o} V_{\tilde{Y}_i^o}^{-1} (\tilde{Y}_i^o - \hat{\mu}_{\tilde{Y}_i^o})$ et de matrice de variance-covariance $V_{\Lambda_i} - C_{\Lambda_i \tilde{Y}_i^o} V_{\tilde{Y}_i^o}^{-1} C_{\Lambda_i \tilde{Y}_i^o}^\top$.

Les prédictions $\hat{\Lambda}_i^C$ spécifiques au sujet i s'écrivent donc :

$$\hat{\Lambda}_i^{(C)} = \mathbb{E} [\Lambda_i | \tilde{Y}_i^o] = \mu_{\Lambda_i} + C_{\Lambda_i \tilde{Y}_i^o} V_{\tilde{Y}_i^o}^{-1} (\tilde{Y}_i^o - \hat{\mu}_{\tilde{Y}_i^o}) \quad (3.35)$$

3.4.2 Prédictions des observations

Les prédictions marginales et conditionnelles des observations peuvent être obtenues analytiquement dans l'échelle des processus latents à partir de la matrice P reliant les processus latents aux observations et respectivement à partir des prédictions marginales et spécifiques aux sujets.

Les prédictions marginales $\mathbf{Y}_i^{(M)}$ des observations et les prédictions spécifiques $\mathbf{Y}_i^{(C)}$ aux sujets des observations peuvent être aussi faites dans l'échelle naturelle des marqueurs comme suit :

$$\mathbf{Y}_i^{(M)} = \mathbb{E} [H^{-1}(\tilde{Y}_i)] = \int_{\mathbb{R}^{K_i}} H^{-1}(u) \phi_{\tilde{Y}_i}(u) du \quad (3.36)$$

$$\mathbf{Y}_i^{(C)} = \mathbb{E} [H^{-1}(\tilde{Y}_i | \tilde{Y}_i^o)] = \int_{\mathbb{R}^{K_i}} H^{-1}(u) \phi_{\tilde{Y}_i | \tilde{Y}_i^o}(u) du. \quad (3.37)$$

avec ϕ la fonction de densité d'un vecteur Gaussien, K_i la taille du vecteur des mesures répétées Y_i de l'individu i , et $\mathbf{H}^{-1}(\tilde{Y}_i) = \left(\left(H_k^{-1}(\tilde{Y}_{ijk})_{k=1,K}^\top \right)_{j \in \tau_i}^\top \right)^\top$, avec H_k la fonction de lien des mesures du marqueur k .

Les prédictions dans l'échelle naturelle des observations ne peuvent pas être calculées analytiquement. Bien que les fonctions de lien H_k soient inversibles, H_k^{-1} n'a pas nécessairement d'expression simple, notamment avec les I-splines. Sauf dans le cas d'une fonction de lien linéaire, l'inverse de H_k est obtenue numériquement comme décrit en section 3.4.2.1. Une fois H^{-1} connue, l'intégrale sur \mathbb{R}^{K_i} peut être calculée numériquement par la méthode de Monte Carlo décrite en section 3.4.2.2.

3.4.2.1 Calcul numérique de l'inverse de la fonction de lien H_k

Lorsque le marqueur Y_k est Gaussien, la fonction de lien H_k est linéaire et on a :

$$\tilde{y}_{rjk} = H_k(y_{rjk}) = \frac{y_{rjk} - \eta_{0k}}{\eta_{1k}} \quad (3.38)$$

$$\Leftrightarrow y_{rjk} = H_k^{-1}(\tilde{y}_{rjk}) = \tilde{y}_{rjk} \eta_{1k} + \eta_{0k} \quad (3.39)$$

MÉTHODOLOGIE

Lorsque le marqueur Y_k est non Gaussien, la fonction de lien H_k est une base de I-splines (M-splines intégrées) et il n'existe pas une forme analytique pour l'inverse H_k^{-1} . Nous avons donc approximé numériquement le calcul de l'antécédent y_{rjk} de \tilde{y}_{rjk} par la fonction H_k^{-1} par la méthode de Newton-Raphson. Le calcul de l'antécédent y_{rjk} de \tilde{y}_{rjk} par la fonction H_k^{-1} peut être assimilé au calcul de racine de la fonction $H_k(y) - \tilde{y}$:

$$\begin{aligned} H^{-1}(\tilde{y}) = y &\Leftrightarrow H(y) = \tilde{y} \\ &\Leftrightarrow H(y) - \tilde{y} = 0, \end{aligned} \tag{3.40}$$

où l'inconnu y dans l'équation (3.40) est la racine de la fonction $f(y) = H(y) - \tilde{y}$ pour \tilde{y} connu.

L'algorithme de Newton-Raphson pour le calcul numérique de la racine d'une fonction f de \mathbb{R}^d à valeurs dans \mathbb{R}^d se présente comme suit :

Algorithme

1. initialiser le vecteur $y : y^{(0)}$
2. À l'itération k :
 - 2.1. Calculer $y^{(k)} = y^{(k-1)} - (\nabla f(y^{(k-1)}))^{-1} f(y^{(k-1)})$
 - 2.2. Évaluer le critère d'arrêt $\|y^{(k)} - y^{(k-1)}\|_2 < \epsilon$
 - Si le critère d'arrêt est atteint alors passer à l'étape (3)
 - Si le critère d'arrêt n'est pas atteint, refaire l'étape 2 en passant à l'itération $k + 1$ de l'étape (2).
3. Renvoyer la solution $y = y^{(k)}$.

Nous avons implémenté et utilisé cet algorithme avec le critère de convergence ϵ fixé à 1.e-9.

3.4.2.2 Approximation numérique du calcul d'intégrale

Nous avons calculé numériquement les intégrales par la méthode de Monte Carlo (MC). La méthode de Monte Carlo approche le calcul de l'intégrale

$$\int_{\mathbb{R}} g(u)\phi(u)du$$

par la somme finie

$$\frac{1}{R} \sum_{r=1}^R g(U_r),$$

où $(U_r)_{r=1,R}$ est un échantillon de taille R de vecteurs \mathbf{U}_r indépendants et identiquement distribués (i.i.d) de fonction de densité ϕ .

Dans notre problème de calcul de prédictions, posons $\hat{\mathbf{y}}_i = \mathbf{Y}_i^{(M)}$ (respectivement $\hat{\mathbf{y}}_i = \mathbf{Y}_i^{(C)}$) avec $\hat{\mathbf{y}}_i = (\hat{y}_{il})_{l=1,K_i}$ un vecteur de taille K_i .

Pour calculer $\hat{\mathbf{y}}_i$ par la méthode de Monte Carlo, on commence par générer un échantillon i.i.d de taille R du vecteur transformé $\tilde{\mathbf{y}}_i$ de fonction de densité $\phi_{\tilde{\mathbf{Y}}_i}$ (respectivement $\phi_{\tilde{\mathbf{Y}}_i|_{\mathbf{Y}_i^o}}$). On obtient ensuite un échantillon i.i.d de taille R du vecteur \mathbf{y}_i en calculant les inverses de l'échantillon i.i.d du vecteur transformé $\tilde{\mathbf{y}}_i$ par la méthode décrite en section 3.4.2.1. L'approximation par Monte Carlo de chaque élément l du vecteur $\hat{\mathbf{y}}_i$ est donnée par :

$$\hat{y}_{il} = \frac{1}{R} \sum_{r=1}^R \hat{y}_{rl}, \quad \text{pour } l = 1, K_i \quad (3.41)$$

3.5 Implémentation du modèle sous R

Nous avons implémenté dans un package R des programmes informatiques pour estimer notre modèle. Ce package dans sa version actuelle est disponible sur GitHub à l'adresse <https://github.com/bachirtadde/CInLPN>. Il combine les langages de programmation R et C++ pour présenter une interface facile d'utilisation pour les utilisateurs du package tout en accélérant les calculs intensifs. Le package R est nommé CInLPN pour « Causal Inference in Latent Processes Network ». Le package contient plusieurs programmes subdivisés en trois groupes.

3.5.1 Programmes de spécification et de pré-traitement

Il s'agit de programmes écrits en R qui permettent à l'utilisateur de spécifier correctement un modèle et d'identifier toutes les composantes du modèle à savoir :

- le nombre de processus latents,
- les marqueurs de chaque processus latent,
- les sous-modèles du modèle structurel que sont le modèle pour le niveau initial et le modèle pour le taux de variation du système de processus latents,
- le modèle d'observation de chaque marqueur,

MÉTHODOLOGIE

- les contraintes sur les paramètres,
- le jeu de données à utiliser pour l'estimation des paramètres,
- et les options personnalisables par l'utilisateur du package.

3.5.1.1 Les arguments de spécification du modèle

L'exemple ci-dessous illustre la spécification d'un modèle par un utilisateur du package `CInLPN`. Cet exemple spécifie un modèle pour deux processus latents Λ_1 et Λ_2 mesurés respectivement par deux marqueurs $Y_{1,1}$, $Y_{2,1}$ et par trois marqueurs $Y_{1,2}$, $Y_{2,2}$, $Y_{3,2}$. Les arguments à définir pour spécifier complètement ce modèle sont regroupés en 7 groupes.

```
mod1 <-
  CInLPN(structural.model =
    list(
      fixed.LPO = ~1 + AGE0 + gender | 1 + AGE0 + gender,
      fixed.DeltaLP = Y1.1 + Y2.1 | Y1.2 + Y2.2 + Y3.2 ~ 1 + t | 1 + t,
      random.DeltaLP = ~1 + t | 1 + t,
      trans.matrix = ~ 1 + gender,
      delta.time = Delta
    ),

    measurement.model =
    list(
      links = c("linear", "4-quant-2", "3-manual-2",
               "5-quant-3", "3-equi-2"),
      knots = list(NULL, NULL, c(0, 12, 20), NULL, NULL)
    ),

    parameters =
    list(paras.ini = paras.ini, Fixed.para.index = indexparaFixeUser,
         Fixed.para.values = paraFixeUser),
    option = list(parallel = T, nproc = 40, print.info = T, epsa = epsa,
                  epsb = epsb, epsd = epsd, univarmaxiter = 25, MCnr=50 ),
    Time="t",
    subject = "id",
    data = data
  )
```

Le premier groupe d'argument `Structural.model` permet de spécifier complètement le modèle structurel. Il regroupe 5 sous-arguments décrits dans le tableau 3.1.

Le deuxième groupe d'argument `measurement.model` contient une liste de deux sous-arguments (décrits en tableau 3.2) qui permettent de définir le modèle d'observation de chaque marqueur.

Le troisième groupe d'argument `parameters` contient une liste de trois sous-arguments que sont :

- `paras.ini` : initialisé par défaut à `NULL`, il permet de spécifier un vecteur initial pour le problème d'optimisation.

TABLE 3.1 – Liste des sous-arguments de l’argument `Structural.model` du package R `CInLPN`.

| argument | Description |
|------------------------------|---|
| <code>fixed.LP0</code> | permet de spécifier un modèle d’effets fixes sur le niveau initial de chaque processus latent |
| <code>fixed.DeltaLP</code> | permet de spécifier un modèle d’effets fixes sur le taux de variation entre deux temps consécutifs de chaque processus latent |
| <code>randoms.DeltaLP</code> | permet de spécifier un modèle d’effets aléatoires sur le taux de variation entre deux temps consécutifs de chaque processus latents |
| <code>trans.matrix</code> | permet de spécifier un modèle pour les coefficients de la matrice d’influences temporelles entre les processus latent |
| <code>delta.time</code> | permet d’indiquer un pas de discrétisation des processus latents |

- `Fixed.param.index` : permet d’indiquer les indices des paramètres qui ne sont pas à estimer.
- `Fixed.param.values` : permet de spécifier la valeur des paramètres à ne pas estimer.

Le quatrième groupe d’argument `option` qui est optionnel permet de personnaliser des options d’optimisation et d’affichage. Il contient les sous-arguments :

- `parallel` : initialisé par défaut à `FALSE`, il permet de spécifier si le problème d’optimisation est fait par une version parallélisée de l’algorithme de Levenberg-Marquardt. Cet argument permet de gagner du temps lors de l’optimisation sur une machine qui possède plusieurs cœurs de calcul.
- `nproc` : initialisé à 1, cet argument permet d’indiquer le nombre de cœurs de travail à utiliser par le programme d’optimisation.
- `print.info` : initialisé à `FALSE`, il permet d’afficher les itérations lors de l’optimisation.
- `epsa`, `epsb`, `epsd` : initialisés tous les trois à `1.e4`, ils permettent de spécifier les critères de convergence du problème de maximisation de la log-vraisemblance.
- `univarmaxiter` : initialisé à 10, il permet de spécifier le nombre maximal d’itération pour chaque estimation de modèle univarié dans la stratégie d’initialisation du vecteur de paramètres du problème de maximisation de la log-vraisemblance.

TABLE 3.2 – Liste des sous-arguments de l'argument `measurement.model` du package R `CInLPN`.

| argument | Description |
|--------------------|---|
| <code>links</code> | permet de spécifier la fonction de lien de chaque marqueur avec le processus latent qu'il mesure. Il est possible de spécifier un lien linéaire (" <code>linear</code> ") pour un marqueur Gaussien ou des liens par combinaison de I-splines à l'aide de trois arguments sous forme d'une chaîne de caractère : " <code>n-type-d</code> " où <code>n</code> indique le nombre de nœuds total, l'argument <code>type</code> prend les valeurs <code>quant</code> , <code>manual</code> , <code>equi</code> pour indiquer respectivement si les nœuds sont placés aux quantiles, manuellement à des points spécifiés par l'utilisateur ou de manière équidistante. L'argument <code>d</code> indique le degré des I-splines. |
| <code>knots</code> | permet de spécifier éventuellement les noeuds. Cet argument prend la valeur <code>NULL</code> ou un vecteur de valeurs réelles. |

- `MCnr` : initialisé à 100, il permet de spécifier le nombre de répliques dans l'approximation du calcul d'intégrale par la méthode de Monte Carlo dans le calcul des prédictions marginales et spécifiques aux sujets.

Les trois derniers arguments du modèle sont `Time`, `subject` et `data` pour spécifier la variable temps, la variable représentant les identifiants des sujets et enfin le jeu de données.

Après la spécification du modèle, d'autres programmes sont implémentés en R pour pré-traiter, formater le jeu de données, créer les matrices de données nécessaires pour chaque compartiment du modèle et initialiser les paramètres du modèle.

L'initialisation des paramètres est faite de manière stratégique, en estimant un modèle univarié pour chaque processus latent. Les paramètres estimés par les modèles univariés sont utilisés pour initialiser les paramètres du modèle complet en initialisant les paramètres d'interdépendance entre les processus latents à zéro.

3.5.2 Programmes d'estimation et de prédiction

Les tâches d'estimation et de prédiction sont coûteuses en temps de calcul. Pour cela, nous avons écrit un programme en `C++` pour le calcul de la log-vraisemblance pour un jeu de paramètres donnés. Ce dernier est appelé par un programme R qui calcule de ma-

nière itérative l'estimateur du maximum de vraisemblance par l'algorithme de Levenberg-Marquardt [Marquardt, 1963]. L'interface de communication `RcppArmadillo` est utilisée pour faciliter les échanges entre le programme R et le programme C++ [Eddelbuettel et Sanderson, 2014].

Le programme d'estimation converge au bout d'un nombre d'itérations inférieur à un nombre maximal d'itérations et suivant les critères de convergence ϵ_a , ϵ_b et ϵ_d (voir section 3.2.4.2) indiqués par l'utilisateur. Après l'optimisation, un programme C++ est appelé pour calculer les prédictions marginales et spécifiques des observations dans l'échelle des processus latents et dans l'échelle naturelle des données.

Les prédictions dans l'échelle naturelle des marqueurs sont faite telle que décrite dans la section 3.4.2.

3.5.3 Intervalle de confiance des coefficients de la matrice d'influences temporelles

Des programmes R sont aussi écrits pour calculer les intervalles de confiance des coefficients de la matrice d'influences temporelles par la delta méthode.

Chaque coefficient $a_{dd'}$ de la matrice d'influences temporelles est modélisé par la régression linéaire donnée en équation 3.2 de la section 3.1.1.

Plus généralement l'expression du coefficient $a_{dd'}$ peut s'écrire comme une fonction g du vecteur de paramètres $\boldsymbol{\alpha}_{dd'}$ associés :

$$a_{dd'} = g(\boldsymbol{\alpha}_{dd'}), \forall (d, d') \in \{1, \dots, D\}^2, \quad (3.42)$$

avec $\boldsymbol{\alpha}_{dd'}$ les paramètres associés au coefficient $a_{dd'}$ et g une fonction continue et dérivable.

Le vecteur $\hat{\boldsymbol{\alpha}}_{dd'}$ a une loi asymptotique normale $\hat{\boldsymbol{\alpha}}_{dd'} \sim \mathcal{N}(\boldsymbol{\alpha}_{dd'}, \mathbf{V}_{\hat{\boldsymbol{\alpha}}_{dd'}})$, avec $\hat{\boldsymbol{\alpha}}_{dd'}$ l'estimateur du maximum de vraisemblance de $\boldsymbol{\alpha}_{dd'}$ et $\hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}_{dd'}}$ l'estimateur de la matrice de variance-covariance $\mathbf{V}_{\hat{\boldsymbol{\alpha}}_{dd'}}$ de $\hat{\boldsymbol{\alpha}}_{dd'}$ obtenue par inversion de la matrice Hessienne.

Par le développement en série de Taylor de la fonction g au voisinage de $\boldsymbol{\alpha}_{dd'}$ on obtient que :

$$\hat{a}_{dd'} = g(\hat{\boldsymbol{\alpha}}_{dd'}) \approx g(\boldsymbol{\alpha}_{dd'}) + (\hat{\boldsymbol{\alpha}}_{dd'} - \boldsymbol{\alpha}_{dd'})^\top \nabla g(\boldsymbol{\alpha}_{dd'}), \quad (3.43)$$

avec $\hat{a}_{dd'}$ l'estimateur du maximum de vraisemblance du coefficient $a_{dd'}$.

L'estimateur $\hat{a}_{dd'}$ a une distribution asymptotique normale :

$$\hat{a}_{dd'} \sim \mathcal{N}\left(g(\boldsymbol{\alpha}_{dd'}), \nabla g(\boldsymbol{\alpha}_{dd'})^\top \mathbf{V}_{\hat{\boldsymbol{\alpha}}_{dd'}} \nabla g(\boldsymbol{\alpha}_{dd'})\right). \quad (3.44)$$

L'estimation de la matrice de variance-covariance de $\hat{a}_{dd'}$ est donnée par :

$$\nabla g(\hat{\boldsymbol{\alpha}}_{dd'})^\top \hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}_{dd'}} \nabla g(\hat{\boldsymbol{\alpha}}_{dd'}).$$

et l'intervalle de confiance de l'estimateur $\hat{a}_{dd'}$ du coefficient $a_{dd'}$ s'écrit :

$$\text{IC}_{95\%}(\hat{a}_{dd'}) = g(\hat{\boldsymbol{\alpha}}_{dd'}) \pm 1.96 \sqrt{\nabla g(\hat{\boldsymbol{\alpha}}_{dd'})^\top \hat{\mathbf{V}}_{\hat{\boldsymbol{\alpha}}_{dd'}} \nabla g(\hat{\boldsymbol{\alpha}}_{dd'})}. \quad (3.45)$$

Ce calcul est intéressant dans le cas où les coefficients de la matrice d'influences temporelles sont ajustés sur des variables explicatives et/ou des fonctions du temps.

3.5.4 Programmes de sortie des résultats

Le troisième groupe de programmes permet de présenter les résultats de l'estimation et des prédictions à l'utilisateur. Ces programmes écrits en R résument le modèle estimé, renseignent les valeurs des paramètres à l'optimum, leur écart-type et la p-valeur issue du test de Wald, ainsi que les valeurs des critères d'information AIC et BIC.

Nous avons enfin écrit sous R des scripts pour représenter graphiquement l'évolution des coefficients de la matrice d'influences temporelles et les prédictions du modèle.

4 Simulations

Dans le chapitre 3 nous avons présenté notre approche de modélisation innovante dans laquelle nous avons défini en temps discret des modèles paramétriques pour le niveau initial de processus latents, pour leur taux de variation et pour leur structure d'influences temporelles. Nous avons aussi défini un modèle d'observation paramétrique qui relie les observations aux processus latents. Ce modèle théorique est très complexe avec l'implémentation de paramètres à plusieurs niveaux de modélisation, les paramètres centraux étant ceux qui définissent les influences temporelles entre les processus.

Ce chapitre présente des études de simulation pour valider le programme d'estimation et la qualité des estimateurs du maximum de vraisemblance des paramètres du modèle. Il vise aussi à étudier certains cas de mauvaise spécification de modèle et l'impact de la discrétisation du temps sur la structure infinitésimale d'influence entre les processus latents. Avant de présenter les résultats des simulations, nous exposons la méthodologie générale adoptée pour les études de simulation.

4.0.1 Méthodologie des études de simulation

La méthodologie d'analyse est la même pour les deux objectifs visés dans ce chapitre. Elle consiste à réaliser pour R répliques les tâches de :

1. génération des données,
2. et d'estimation de modèle sur les données générées.

4.0.1.1 Génération des données

Les données de chaque réplique r sont générées en simulant pour n sujets les observations de K marqueurs longitudinaux Gaussiens de K processus latents. Les processus latents Gaussiens sont d'abord simulés à partir du modèle structurel des processus qui décrit leur évolution et leur structure d'influences temporelles telles que présentées dans le chapitre 3. La simulation des processus latents passe d'abord par la génération des variables explicatives et des effets aléatoires du modèle structurel. Une fois les processus

SIMULATIONS

latents du sujet i simulés, les marqueurs Y_{ijk} du sujet i sont simulés aux temps d'observation $t_{ijk} \in \tau_i = \{0, \delta, \dots, \delta J\}$ en rajoutant une erreur de mesure Gaussienne au processus latent mesuré par chaque marqueur.

La procédure de génération des données est décrite comme suit :

1. On simule les effets aléatoires $(\mathbf{u}_i, \mathbf{v}_i)^\top$ à partir d'une loi normale multivariée centrée $\mathcal{N}(0, \mathbf{B})$ avec $\mathbf{B} = \mathbf{L}\mathbf{L}^\top$. La matrice de Cholesky L est paramétrée pour garantir que la matrice \mathbf{B} soit définie positive.
2. On simule les variables explicatives C_1 suivant une loi normale $\mathcal{N}(0, 0.64)$ et C_2 suivant une loi de Bernoulli de paramètre $p = 0.37$. Ces variables sont simulées pour mimer respectivement dans l'échantillon de la base de données ADNI 1 qui sera utilisée dans les applications, la variable explicative continue « âge à l'inclusion » centrée sur 75.4 ans et la variable explicative binaire « MCI » qui indique si le sujet présente de légers troubles cognitifs ou pas.
3. On définit les régresseurs $\mathbf{X}_i^{(0)}$, $\mathbf{X}_i(t_{ij})$ et $\mathbf{Z}_i(t_{ij})$ suivant le scénario considéré à partir du temps et des variables explicatives C_1 et C_2 .
4. On simule $\mathbf{\Lambda}_i(0) = (\Lambda_{i1}(0), \dots, \Lambda_{iK}(0))^\top = \mathbf{X}_i^{(0)}\boldsymbol{\beta} + \mathbf{u}_i$, puis de façon incrémentale

$$\mathbf{\Lambda}_i(t_{ij}) = \delta(\mathbf{X}_i(t_{ij})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{ij})\mathbf{v}_i) + (\mathbb{I}_D + \delta\mathbf{\Lambda}_i(t_{ij} - \delta)),$$

pour $t_{ij} = \delta, \dots, \delta J$ avec δ le pas de discrétisation.

5. Pour chaque marqueur Y_{ijk} on simule ϵ_{ijk} suivant une loi normale $\mathcal{N}(0, \sigma_k^2)$, ensuite on calcule :

$$Y_{ijk} = \eta_{0k} + \eta_{1k}\Lambda_{id}(t_{ijk}) + \epsilon_{ijk} \quad (4.1)$$

Les paramètres $\boldsymbol{\theta} = (\text{vec}(\mathbf{L}), \boldsymbol{\beta}, \boldsymbol{\gamma}, (\eta_{0k}, \eta_{1k})^\top, (\sigma_k)_{k=1, \dots, K})^\top$ sont choisis à partir de l'estimation du modèle de génération des données sur les données ADNI 1.

4.0.2 Validation du programme d'estimation

Pour valider le programme d'estimation, nous avons simulé $R = 100$ répliques d'un jeu de données de $N = 512$ sujets. Les données de chaque sujet sont les observations aux temps $t_{ij} \in \tau_i = \{0, 1, 2, 3, 4, 5, J = 6\}$ de K marqueurs de K processus latents.

Scénario 1

Dans le premier cas d'étude de simulation, nous avons considéré que $K = 3$ processus latents mesurés par $K = 3$ marqueurs longitudinaux Gaussiens Y_{ijk} , $k = 1, 2, 3$. Les

processus latents ont un taux de variation qui a une évolution linéaire du temps entre les temps discrets t et $t + 1$. Les processus sont liés entre eux par une matrice d'influences temporelles A constante au cours du temps mais qui dépend de la variable explicative C_2 . Le modèle d'évolution des processus latents et de leur structure d'influences temporelles est défini comme suit :

$$\left\{ \begin{array}{l} \Lambda_{i1}(0) = \beta_{01} + \beta_{11}C_{1,i} + \beta_{21}C_{2,i} + u_{i1} \\ \Lambda_{i2}(0) = \beta_{02} + \beta_{12}C_{1,i} + \beta_{22}C_{2,i} + u_{i2} \\ \Lambda_{i3}(0) = \beta_{03} + \beta_{13}C_{1,i} + \beta_{23}C_{2,i} + u_{i3} \\ \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}C_{2,i} + \gamma_{21}t_{ij} + \gamma_{31}C_{2,i}t_{ij} + v_{i1} \\ \quad + \alpha_{11}^0\Lambda_{i1}(t_{ij} - \delta) + (\alpha_{12}^0 + \alpha_{12}^1C_{2,i})\Lambda_{i2}(t_{ij} - \delta) + (\alpha_{13}^0 \\ \quad + \alpha_{13}^1C_{2,i})\Lambda_{i3}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}C_{2,i} + \gamma_{22}t_{ij} + \gamma_{32}C_{2,i}t_{ij} + v_{i2} \\ \quad + (\alpha_{21}^0 + \alpha_{21}^1C_{2,i})\Lambda_{i1}(t_{ij} - \delta) + \alpha_{22}^0\Lambda_{i2}(t_{ij} - \delta) \\ \quad + (\alpha_{23}^0 + \alpha_{23}^1C_{2,i})\Lambda_{i3}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i3}(t_{ij})}{\delta} = \gamma_{03} + \gamma_{13}C_{2,i} + \gamma_{23}t_{ij} + \gamma_{33}C_{2,i}t_{ij} + v_{i3} \\ \quad + (\alpha_{31}^0 + \alpha_{31}^1C_{2,i})\Lambda_{i1}(t_{ij} - \delta) + (\alpha_{32}^0 + \alpha_{32}^1)\Lambda_{i2}(t_{ij} - \delta) \\ \quad + \alpha_{33}^0\Lambda_{i3}(t_{ij} - \delta) \end{array} \right.$$

pour $t_{ij} \in \tau_i = \{1, 2, 3, 4, 5, J = 6\}$, $\Delta\Lambda_{id} = \Lambda_{id}(t_{ij}) - \Lambda_{id}(t_{ij} - \delta)$, pour $d = 1, 2, 3$ et $\delta = 1$, β , α et γ des paramètres d'effets fixes, $(u_{i1}, v_{i1}, u_{i2}, v_{i2}, u_{i3}, v_{i3})^\top \sim \mathcal{N}(0, \mathbf{B})$.

Les équations d'observation pour chaque marqueur Y_{ijk} , $k = 1, 2, 3$ sont celles données en (4.1).

Les résultats de cette étude de simulations sont présentés dans le tableau 4.1. Le modèle comporte 36 paramètres à estimer. Tous les paramètres sont bien estimés avec des biais relatifs et des taux de couverture à 95% corrects. Les écart-types des paramètres sont aussi bien estimés. Notons que sur les 100 répliques de données, le programme d'estimation à convergé pour 96 répliques.

SIMULATIONS

TABLE 4.1 – Étude de validation par simulation du programme d'estimation (scénario 1) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture à 95% de 96 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|---------------|----------------|--------------------|---------------|--------------|-----------------------|--------------------|
| β_{11} | -0.627 | -0.601 | 4.1 | 0.135 | 0.071 | 93.7 |
| β_{21} | -0.841 | -0.836 | 0.5 | 0.098 | 0.097 | 96.8 |
| β_{12} | -0.317 | -0.302 | 4.8 | 0.062 | 0.056 | 94.7 |
| β_{22} | -1.801 | -1.808 | 0.4 | 0.122 | 0.121 | 91.6 |
| β_{13} | 0.001 | 0.005 | 418.3 | 0.076 | 0.080 | 97.9 |
| β_{23} | -1.886 | -1.916 | 1.6 | 0.127 | 0.143 | 95.8 |
| γ_{01} | -0.084 | -0.085 | 1.7 | 0.013 | 0.015 | 97.9 |
| γ_{11} | -0.037 | -0.031 | 14.9 | 0.036 | 0.021 | 88.5 |
| γ_{21} | 0.007 | 0.008 | 7.9 | 0.004 | 0.004 | 92.7 |
| γ_{31} | 0.004 | 0.004 | 0.3 | 0.004 | 0.005 | 95.8 |
| γ_{02} | 0.119 | 0.122 | 2.4 | 0.028 | 0.031 | 96.9 |
| γ_{12} | -0.443 | -0.443 | 0.1 | 0.060 | 0.066 | 96.9 |
| γ_{22} | -0.014 | -0.015 | 3.7 | 0.006 | 0.007 | 93.7 |
| γ_{32} | 0.001 | 0.001 | 21.1 | 0.009 | 0.009 | 94.8 |
| γ_{03} | 0.027 | 0.024 | 12.0 | 0.052 | 0.060 | 95.8 |
| γ_{13} | -0.171 | -0.169 | 1.3 | 0.069 | 0.076 | 95.8 |
| γ_{23} | -0.021 | -0.021 | 0.2 | 0.012 | 0.014 | 93.7 |
| γ_{33} | -0.005 | -0.004 | 13.2 | 0.016 | 0.018 | 92.7 |
| $L.u_2u_1$ | 0.340 | 0.337 | 0.9 | 0.050 | 0.048 | 95.8 |
| $L.u_3u_1$ | 0.179 | 0.181 | 1.1 | 0.062 | 0.054 | 92.7 |
| $L.v_1u_1$ | 0.045 | 0.049 | 9.9 | 0.026 | 0.013 | 93.7 |
| $L.u_3u_2$ | 0.301 | 0.310 | 3.1 | 0.059 | 0.056 | 93.7 |
| $L.v_2u_2$ | 0.320 | 0.321 | 0.3 | 0.045 | 0.041 | 88.5 |
| $L.v_3u_3$ | 0.122 | 0.122 | 0.1 | 0.041 | 0.038 | 94.8 |
| $L.v_1$ | 0.058 | 0.057 | 1.7 | 0.007 | 0.004 | 92.7 |
| $L.v_2$ | 0.177 | 0.175 | 1.2 | 0.019 | 0.019 | 95.8 |
| $L.v_3$ | 0.235 | 0.235 | 0.1 | 0.027 | 0.027 | 94.8 |
| a_{11}^0 | -0.045 | -0.038 | 15.9 | 0.041 | 0.013 | 94.8 |

Suite de la TABLE 4.1 - Étude de validation par simulation du programme d'estimation (scénario 1) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 96 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|-------------|-------------------|-----------------------|------------------|-----------------|--------------------------|-----------------------|
| a_{12}^0 | 0.003 | 0.003 | 5.0 | 0.005 | 0.005 | 92.7 |
| a_{12}^1 | 0.025 | 0.024 | 3.2 | 0.007 | 0.007 | 90.6 |
| a_{13}^0 | 0.005 | 0.006 | 14.4 | 0.005 | 0.005 | 92.7 |
| a_{13}^1 | -0.004 | -0.004 | 4.0 | 0.006 | 0.007 | 97.9 |
| a_{21}^0 | 0.148 | 0.146 | 1.6 | 0.029 | 0.027 | 90.6 |
| a_{21}^1 | 0.024 | 0.028 | 15.6 | 0.024 | 0.022 | 92.7 |
| a_{22}^0 | -0.301 | -0.300 | 0.2 | 0.043 | 0.040 | 90.6 |
| a_{23}^0 | 0.064 | 0.064 | 0.6 | 0.017 | 0.016 | 93.7 |
| a_{23}^1 | 0.011 | 0.010 | 9.3 | 0.019 | 0.017 | 90.6 |
| a_{31}^0 | 0.053 | 0.048 | 9.5 | 0.022 | 0.026 | 97.9 |
| a_{31}^1 | 0.041 | 0.045 | 9.7 | 0.029 | 0.034 | 92.7 |
| a_{32}^0 | 0.049 | 0.054 | 10.3 | 0.023 | 0.024 | 93.7 |
| a_{32}^1 | 0.086 | 0.082 | 4.1 | 0.024 | 0.026 | 96.9 |
| a_{33}^0 | -0.136 | -0.133 | 2.2 | 0.043 | 0.041 | 93.7 |
| σ_1 | 0.235 | 0.236 | 0.6 | 0.010 | 0.008 | 91.7 |
| σ_2 | 0.404 | 0.405 | 0.1 | 0.016 | 0.016 | 94.8 |
| σ_3 | 0.732 | 0.740 | 1.1 | 0.031 | 0.034 | 96.9 |
| η_{01} | 3.193 | 3.175 | 0.6 | 0.283 | 0.310 | 96.9 |
| η_{11} | 4.197 | 4.181 | 0.4 | 0.159 | 0.135 | 89.6 |
| η_{02} | 3.821 | 3.826 | 0.1 | 0.200 | 0.200 | 95.8 |
| η_{12} | 2.466 | 2.464 | 0.1 | 0.088 | 0.086 | 94.8 |
| η_{03} | 2.501 | 2.499 | 0.1 | 0.114 | 0.129 | 97.9 |
| η_{13} | 1.376 | 1.359 | 1.2 | 0.054 | 0.056 | 95.8 |

Scénario 2

Dans un deuxième scénario de simulation, nous avons évalué le programme d'estimation en présence de données manquantes. Pour cela, nous avons repris le cas d'étude 1 avec des données manquantes générées pour mimer les données manquantes dans l'échantillon

SIMULATIONS

de jeu de données ADNI 1. Dans l'échantillon de jeu de données ADNI1, le taux de visites manquantes est de 0.26. Nous avons donc tiré sans remise $0.26 \times (J+1) \times 512 \approx 932$ visites qui seront manquantes. De plus, à chaque visite non manquante un des trois marqueurs peut être manquant avec une probabilité $p = 0.06$.

Le tableau 4.2 contient les résultats des simulations. On voit qu'à nouveau, tous les paramètres du modèles et leur écart-type sont bien estimés avec des biais et taux de couvertures corrects.

TABLE 4.2 – Étude de validation par simulation du programme d'estimation en présence de données manquantes (scénario 2) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 96 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1 avec environ 25% de taux de visites manquantes et une probabilité de 0.05 qu'au moins un marqueur soit manquant à une visite effectuée.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|---------------|----------------|--------------------|---------------|--------------|-----------------------|--------------------|
| β_{11} | -0.627 | -0.606 | 3.4 | 0.114 | 0.072 | 92.7 |
| β_{21} | -0.841 | -0.839 | 0.2 | 0.095 | 0.099 | 96.9 |
| β_{12} | -0.317 | -0.308 | 3.0 | 0.057 | 0.061 | 96.9 |
| β_{22} | -1.801 | -1.819 | 0.9 | 0.127 | 0.126 | 93.7 |
| β_{13} | 0.001 | 0.003 | 210.3 | 0.081 | 0.088 | 98.9 |
| β_{23} | -1.886 | -1.929 | 2.3 | 0.145 | 0.154 | 95.8 |
| γ_{01} | -0.084 | -0.083 | 0.6 | 0.017 | 0.018 | 94.8 |
| γ_{11} | -0.037 | -0.035 | 4.9 | 0.034 | 0.025 | 93.7 |
| γ_{21} | 0.007 | 0.007 | 1.6 | 0.005 | 0.005 | 94.8 |
| γ_{r1} | 0.004 | 0.005 | 27.7 | 0.007 | 0.006 | 92.7 |
| γ_{02} | 0.119 | 0.122 | 2.5 | 0.033 | 0.037 | 96.9 |
| γ_{12} | -0.443 | -0.449 | 1.4 | 0.074 | 0.078 | 94.8 |
| γ_{22} | -0.014 | -0.014 | 0.5 | 0.007 | 0.008 | 97.9 |
| γ_{32} | 0.001 | -0.000 | 117.2 | 0.010 | 0.011 | 93.7 |
| γ_{03} | 0.027 | 0.021 | 23.9 | 0.066 | 0.059 | 91.7 |
| γ_{13} | -0.171 | -0.168 | 1.8 | 0.083 | 0.082 | 93.7 |
| γ_{23} | -0.021 | -0.020 | 3.6 | 0.015 | 0.014 | 93.7 |
| γ_{33} | -0.005 | -0.005 | 4.6 | 0.020 | 0.020 | 95.8 |
| $L.u_2u_1$ | 0.340 | 0.336 | 1.3 | 0.049 | 0.050 | 97.9 |

Suite de la TABLE 4.2 - Étude de validation par simulation du programme d'estimation en présence de données manquantes (scénario 2 avec 25% de taux de visites manquantes et une probabilité de 0.05 qu'au moins un marqueur soit manquant à une visite effectuée.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|-------------|-------------------|-----------------------|------------------|-----------------|--------------------------|-----------------------|
| $L.u_3u_1$ | 0.179 | 0.180 | 0.8 | 0.068 | 0.058 | 90.6 |
| $L.v_1u_1$ | 0.045 | 0.050 | 10.2 | 0.033 | 0.015 | 90.6 |
| $L.u_3u_2$ | 0.301 | 0.313 | 4.1 | 0.066 | 0.062 | 93.7 |
| $L.v_2u_2$ | 0.320 | 0.325 | 1.6 | 0.058 | 0.052 | 92.7 |
| $L.v_3u_3$ | 0.122 | 0.122 | 0.1 | 0.050 | 0.047 | 94.8 |
| $L.v_1$ | 0.058 | 0.058 | 0.8 | 0.007 | 0.005 | 95.8 |
| $L.v_2$ | 0.177 | 0.177 | 0.2 | 0.023 | 0.024 | 96.9 |
| $L.v_3$ | 0.235 | 0.237 | 0.9 | 0.034 | 0.033 | 90.6 |
| a_{11}^0 | -0.045 | -0.039 | 13.8 | 0.045 | 0.015 | 91.7 |
| a_{12}^0 | 0.003 | 0.003 | 15.5 | 0.006 | 0.006 | 93.7 |
| a_{12}^1 | 0.025 | 0.024 | 5.4 | 0.008 | 0.008 | 90.6 |
| a_{13}^0 | 0.005 | 0.005 | 0.1 | 0.006 | 0.006 | 90.6 |
| a_{13}^1 | -0.004 | -0.003 | 18.2 | 0.007 | 0.008 | 94.8 |
| a_{21}^0 | 0.148 | 0.147 | 0.5 | 0.033 | 0.032 | 92.7 |
| a_{21}^1 | 0.024 | 0.027 | 11.3 | 0.024 | 0.024 | 95.8 |
| a_{22}^0 | -0.301 | -0.303 | 0.7 | 0.055 | 0.051 | 91.7 |
| a_{23}^0 | 0.064 | 0.064 | 0.5 | 0.021 | 0.019 | 95.8 |
| a_{23}^1 | 0.011 | 0.010 | 11.0 | 0.021 | 0.020 | 94.8 |
| a_{31}^0 | 0.053 | 0.047 | 10.9 | 0.024 | 0.025 | 95.8 |
| a_{31}^1 | 0.041 | 0.047 | 14.5 | 0.034 | 0.031 | 94.8 |
| a_{32}^0 | 0.049 | 0.055 | 12.5 | 0.027 | 0.027 | 91.7 |
| a_{32}^1 | 0.086 | 0.082 | 5.1 | 0.027 | 0.029 | 96.9 |
| a_{33}^0 | -0.136 | -0.135 | 0.8 | 0.055 | 0.051 | 93.7 |
| σ_1 | 0.235 | 0.237 | 0.7 | 0.009 | 0.009 | 93.7 |
| σ_2 | 0.404 | 0.406 | 0.5 | 0.018 | 0.017 | 94.8 |
| σ_3 | 0.732 | 0.741 | 1.2 | 0.036 | 0.039 | 96.9 |
| η_{01} | 3.193 | 3.179 | 0.4 | 0.286 | 0.313 | 95.8 |
| η_{11} | 4.197 | 4.170 | 0.6 | 0.139 | 0.136 | 92.7 |
| η_{02} | 3.821 | 3.837 | 0.4 | 0.200 | 0.206 | 93.7 |

SIMULATIONS

Suite de la TABLE 4.2 - Étude de validation par simulation du programme d'estimation en présence de données manquantes (scénario 2 avec 25% de taux de visites manquantes et une probabilité de 0.05 qu'au moins un marqueur soit manquant à une visite effectuée.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|-------------|-------------------|-----------------------|------------------|-----------------|--------------------------|-----------------------|
| η_{12} | 2.466 | 2.454 | 0.5 | 0.098 | 0.092 | 93.7 |
| η_{03} | 2.501 | 2.510 | 0.3 | 0.120 | 0.133 | 96.9 |
| η_{13} | 1.376 | 1.358 | 1.3 | 0.063 | 0.064 | 95.8 |

Scénarios 3 et 4

Le cas 3 d'étude de simulation évalue le programme d'estimation dans le cas d'une structure d'influences temporelles qui évolue au cours du temps en l'absence de données manquantes (scénario 3) et en présence de données manquantes (scénario 4). Nous avons ici deux processus latents Gaussiens dont le taux de variation au cours du temps est linéaire dans le temps et les influences temporelles entre les processus sont modélisées par une base de B-splines quadratiques avec un nœud interne placé à la médiane =3 des temps de visite $(S_l(t))_{l=1,2,3}$. Le modèle de génération des données est défini comme suit :

$$\begin{cases}
 \Lambda_{i1}(0) &= \beta_{01} + \beta_{11}C_{2,i} + u_{i1} \\
 \Lambda_{i2}(0) &= \beta_{02} + \beta_{12}C_{2,i} + u_{i2} \\
 \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} &= \gamma_{01} + \gamma_{11} + v_{i1} + a_{11}(t_{ij} - \delta)\Lambda_{i1}(t_{ij} - \delta) + a_{12}(t_{ij} - \delta)\Lambda_{i2}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} &= \gamma_{02} + \gamma_{12} + v_{i2} + a_{21}(t_{ij} - \delta)\Lambda_{i1}(t_{ij} - \delta) + a_{22}(t_{ij} - \delta)\Lambda_{i2}(t_{ij} - \delta)
 \end{cases}$$

avec $t_{ij} \in \tau_i = \{1, 2, 3, 4, 5, J = 6\}$,

$$(u_{i1}, v_{i1}, u_{i2}, v_{i2})^\top \sim \mathcal{N}(0, \mathbf{B}),$$

$$a_{dd'}(t) = \alpha_{dd'}^0 + \alpha_{dd'}^1 S_1(t) + \alpha_{dd'}^2 S_2(t) + \alpha_{dd'}^3 S_3(t), \forall d, d' \in \{1, 2\}^2. \quad (4.2)$$

Les résultats sont résumés dans le tableau 4.3. Ici nous avons réalisé $R = 200$ répliques. Les résultats des estimations sans données manquantes (scénario 3) sont présentés à gauche et les résultats de simulations avec des données manquantes (scénario 4) à droite (taux de visites manquantes de 15% et taux de mesure manquante à un marqueur de 7%). Tous les paramètres et leur écart-type sont bien estimés avec des biais et des taux de couvertures corrects dans les deux cas de simulations (avec ou sans données manquantes).

TABLE 4.3 – Étude de validation par simulation du programme d'estimation pour un modèle présentant une structure d'influences temporelles entre les processus qui dépend du temps (scénario 3 à gauche sans données manquantes, scénario 4 à droite avec données manquantes) : Estimation moyenne ($\hat{\theta}$), écart-type empirique (ETE), écart-type asymptotique (ETA) et taux de couverture à 95% (TC) de 200 répliques d'un échantillon de 512 sujets.

| | θ | sans données manquantes | | | | avec données manquantes * | | | |
|-----------------|----------|-------------------------|-------|-------|-------|---------------------------|-------|-------|-------|
| | | $\hat{\theta}$ | ETE | ETA | TC(%) | $\hat{\theta}$ | ETE | ETA | TC(%) |
| β_{01} | -1.661 | -1.655 | 0.104 | 0.112 | 96.0 | -1.653 | 0.111 | 0.113 | 98.0 |
| β_{02} | -1.803 | -1.831 | 0.146 | 0.128 | 92.0 | -1.834 | 0.150 | 0.133 | 90.0 |
| γ_{01} | 0.278 | 0.281 | 0.043 | 0.046 | 96.0 | 0.285 | 0.045 | 0.050 | 98.0 |
| γ_{11} | -0.152 | -0.154 | 0.023 | 0.023 | 96.0 | -0.156 | 0.024 | 0.026 | 96.0 |
| γ_{02} | 0.117 | 0.121 | 0.080 | 0.080 | 94.0 | 0.111 | 0.093 | 0.086 | 91.0 |
| γ_{12} | -0.143 | -0.145 | 0.040 | 0.041 | 96.0 | -0.140 | 0.046 | 0.045 | 95.0 |
| $L.v_1u_1$ | 0.041 | 0.040 | 0.019 | 0.022 | 97.0 | 0.039 | 0.020 | 0.024 | 98.9 |
| $L.v_2u_2$ | 0.012 | 0.026 | 0.021 | 0.039 | 97.0 | 0.029 | 0.025 | 0.044 | 97.0 |
| $L.v_1$ | 0.198 | 0.197 | 0.015 | 0.014 | 94.0 | 0.196 | 0.016 | 0.014 | 92.0 |
| $L.v_2$ | 0.373 | 0.372 | 0.027 | 0.027 | 95.0 | 0.371 | 0.029 | 0.029 | 93.0 |
| α_{11}^0 | 0.112 | 0.109 | 0.040 | 0.044 | 96.0 | 0.107 | 0.046 | 0.048 | 97.0 |
| α_{11}^1 | 0.003 | 0.010 | 0.047 | 0.050 | 95.0 | 0.015 | 0.052 | 0.056 | 96.0 |
| α_{11}^2 | -0.033 | -0.050 | 0.070 | 0.070 | 97.0 | -0.056 | 0.070 | 0.079 | 98.0 |
| α_{11}^3 | -0.132 | -0.117 | 0.095 | 0.093 | 95.0 | -0.108 | 0.099 | 0.104 | 97.0 |
| α_{12}^0 | 0.168 | 0.170 | 0.039 | 0.037 | 93.0 | 0.172 | 0.044 | 0.041 | 92.0 |
| α_{12}^1 | 0.079 | 0.067 | 0.043 | 0.043 | 92.0 | 0.067 | 0.049 | 0.048 | 92.0 |
| α_{12}^2 | 0.143 | 0.160 | 0.062 | 0.057 | 92.0 | 0.161 | 0.065 | 0.064 | 92.0 |
| α_{12}^3 | 0.070 | 0.050 | 0.080 | 0.073 | 90.0 | 0.048 | 0.088 | 0.081 | 89.0 |
| α_{21}^0 | 0.238 | 0.225 | 0.071 | 0.068 | 95.0 | 0.220 | 0.076 | 0.073 | 92.0 |
| α_{21}^1 | 0.138 | 0.159 | 0.092 | 0.090 | 94.0 | 0.162 | 0.097 | 0.097 | 93.0 |
| α_{21}^2 | 0.359 | 0.331 | 0.128 | 0.126 | 94.0 | 0.330 | 0.134 | 0.135 | 95.0 |
| α_{21}^3 | 0.014 | 0.048 | 0.170 | 0.165 | 94.0 | 0.045 | 0.179 | 0.178 | 95.0 |
| α_{22}^1 | 0.108 | 0.108 | 0.077 | 0.074 | 91.0 | 0.102 | 0.085 | 0.081 | 89.0 |
| α_{22}^2 | 0.037 | 0.024 | 0.082 | 0.084 | 97.0 | 0.025 | 0.091 | 0.088 | 95.0 |
| α_{22}^3 | -0.067 | -0.046 | 0.105 | 0.107 | 90.0 | -0.049 | 0.116 | 0.113 | 91.0 |
| α_{22}^4 | -0.066 | -0.093 | 0.133 | 0.133 | 96.0 | -0.085 | 0.154 | 0.142 | 94.0 |

SIMULATIONS

Suite de la TABLE 4.3 - Étude de validation par simulation du programme d'estimation pour un modèle présentant une structure d'influences temporelles entre les processus qui dépend du temps (scénario 3 à gauche sans données manquantes, scénario 4 à droite avec données manquantes).

| | θ | sans données manquantes | | | | avec données manquantes * | | | |
|-------------|----------|-------------------------|-------|-------|-------|---------------------------|-------|-------|-------|
| | | $\hat{\theta}$ | ETE | ETA | TC(%) | $\hat{\theta}$ | ETE | ETA | TC(%) |
| σ_1 | 0.387 | 0.391 | 0.015 | 0.015 | 93.0 | 0.391 | 0.015 | 0.016 | 96.0 |
| σ_2 | 0.713 | 0.713 | 0.035 | 0.033 | 94.0 | 0.709 | 0.035 | 0.035 | 95.0 |
| η_{01} | 3.769 | 3.725 | 0.192 | 0.197 | 95.0 | 3.719 | 0.199 | 0.199 | 93.0 |
| η_{11} | 2.489 | 2.511 | 0.138 | 0.120 | 92.0 | 2.515 | 0.144 | 0.124 | 91.0 |
| η_{02} | 2.580 | 2.561 | 0.092 | 0.090 | 95.0 | 2.563 | 0.094 | 0.092 | 95.0 |
| η_{12} | 1.412 | 1.413 | 0.063 | 0.060 | 93.0 | 1.415 | 0.063 | 0.064 | 95.0 |

* 15% taux de visites manquantes, 7% de mesures manquantes à une visite.

Conclusion

Cette première partie de l'étude valide les bonnes propriétés des estimateurs du maximum de vraisemblance des paramètres du modèle lorsque le modèle est bien spécifié, et ce pour un nombre de sujets relativement bas (env. 500), un nombre de mesures répétées raisonnable (maximum 7), un taux de données manquantes relativement élevé (25%) et selon plusieurs spécifications de la matrice d'influences temporelles. Ces résultats confirment ceux trouvés dans les simulations du papier méthodologique qui n'incluaient que deux processus (voir annexe).

4.0.3 Robustesse du modèle à une mauvaise spécification

Dans la section précédente, nous avons considéré le cas bien spécifié : le modèle estimé était le même modèle que celui utilisé pour simuler les données.

Dans la présente section, nous avons évalué le comportement des estimateurs du maximum de vraisemblance dans deux cas de mauvaise spécification de la trajectoire des processus :

- mauvaise spécification de la trajectoire moyenne des processus (scénarios 5 et 6),
- mauvaise spécification des corrélations résiduelles entre les niveaux initiaux des processus (scénario 7)

Scénario 5 : Mauvaise spécification de l'effet de variable explicative sur la trajectoire de processus

Dans ce cas d'étude de simulation, nous avons évalué le comportement des estimateurs du maximum de vraisemblance dans le cas d'une mauvaise spécification des trajectoires des processus latents qui dépendent d'une variable explicative binaire C_2 . Nous avons simulé le modèle incluant les associations avec C_2 puis nous avons estimé un modèle sans ajuster les trajectoires des processus latents sur la variable explicative C_2 . Le modèle de génération des données est décrit comme suit :

$$\begin{cases} \Lambda_{i1}(0) = \beta_{01} + \beta_{11}C_{2,i} + u_{i1} \\ \Lambda_{i2}(0) = \beta_{02} + \beta_{12}C_{2,i} + u_{i2} \\ \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}C_{2,i} + v_{i1} + a_{11}\Lambda_{i1}(t_{ij} - \delta) + a_{12}\Lambda_{i2}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}C_{2,i} + v_{i2} + a_{21}\Lambda_{i1}(t_{ij} - \delta) + a_{22}\Lambda_{i2}(t_{ij} - \delta) \end{cases} \quad (4.3)$$

avec $t_{ij} \in \tau_i = \{1, 2, 3, 4, 5, J = 6\}$, $(u_{i1}, v_{i1}, u_{i2}, v_{i2})^\top \sim \mathcal{N}(0, \mathbf{B})$.

Les résultats sont présentés dans le tableau 4.4. Comme attendu, une mauvaise spécification des trajectoires des processus entraîne des biais considérables sur les estimations des effets fixes des trajectoires mais cela impacte aussi dans une moindre mesure la structure d'influence temporelle. En effet, si les influences temporelles vers le processus 1 sont correctement estimées, celles vers le processus 2 ont un biais non négligeable allant jusqu'à 22% dans ce scénario. Notons que dans cette simulation, l'effet simulé de la variable explicative C_2 est majeur sur le niveau initial tout comme sur le taux de variation des processus.

TABLE 4.4 – Résultats des simulations en cas de mauvaise spécification des trajectoires des processus (scénario 5) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|---------------|----------------|--------------------|---------------|--------------|-----------------------|--------------------|
| γ_{01} | -0.123 | -0.047 | 61.7 | 0.008 | 0.009 | 0.0 |
| γ_{02} | -0.133 | -0.148 | 11.6 | 0.014 | 0.015 | 84.0 |
| $L.u_1v_1$ | 0.180 | 0.182 | 0.9 | 0.028 | 0.030 | 98.0 |
| $L.u_2v_2$ | 0.045 | 0.076 | 68.5 | 0.025 | 0.024 | 78.0 |
| $l.u_1$ | 0.146 | 0.119 | 18.9 | 0.010 | 0.011 | 31.0 |

SIMULATIONS

Suite de la TABLE 4.4 - Résultats des simulations en cas de mauvaise spécification des trajectoires des processus (scénario 5) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|-------------|----------------|--------------------|---------------|--------------|-----------------------|--------------------|
| $L.u_2$ | 0.198 | 0.163 | 18.0 | 0.013 | 0.014 | 27.0 |
| a_{11} | -0.156 | -0.169 | 8.4 | 0.029 | 0.030 | 93.0 |
| a_{12} | 0.089 | 0.092 | 2.9 | 0.011 | 0.011 | 93.0 |
| a_{21} | 0.109 | 0.134 | 22.7 | 0.013 | 0.014 | 66.0 |
| a_{22} | -0.055 | -0.094 | 70.9 | 0.026 | 0.026 | 70.0 |
| σ_1 | 0.392 | 0.307 | 21.6 | 0.011 | 0.011 | 0.0 |
| σ_2 | 0.683 | 0.512 | 24.9 | 0.018 | 0.020 | 0.0 |
| η_{01} | 2.714 | 1.082 | 60.1 | 0.168 | 0.151 | 0.0 |
| η_{11} | 1.456 | 0.938 | 35.6 | 0.092 | 0.094 | 0.0 |
| η_{02} | 1.609 | 1.820 | 13.1 | 0.029 | 0.030 | 0.0 |
| η_{12} | 1.216 | 1.403 | 15.4 | 0.022 | 0.025 | 0.0 |

Scénario 6 : Mauvaise spécification de la forme de la trajectoire de processus

Dans cette étude de simulation nous avons investigué l'effet d'une mauvaise spécification du taux de variation des trajectoires des processus latents sur les estimateurs du maximum de vraisemblance des paramètres du modèle. Pour cela, nous avons généré des données à partir d'un modèle dans lequel le taux de variation des processus est une fonction linéaire du temps t , puis nous avons estimé un modèle avec des taux de variation constants sur les données simulées. Le modèle de génération des données est décrit comme suit :

$$\left\{ \begin{array}{l} \Lambda_{i1}(0) = \beta_{01} + u_{i1} \\ \Lambda_{i2}(0) = \beta_{02} + u_{i2} \\ \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}t_{ij} + v_{i1} + a_{11}\Lambda_{i1}(t_{ij} - \delta) + a_{12}\Lambda_{i2}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}t_{ij} + v_{i2} + a_{21}\Lambda_{i1}(t_{ij} - \delta) + a_{22}\Lambda_{i2}(t_{ij} - \delta) \end{array} \right. \quad (4.4)$$

avec $t_{ij} \in \tau_i = \{1, 2, 3, 4, 5, J = 6\}$, $(u_{i1}, v_{i1}, u_{i2}, v_{i2})^\top \sim \mathcal{N}(0, \mathbf{B})$.

Les résultats de cette étude de simulations sont présentés dans le tableau 4.5. Une mauvaise spécification de la forme d'évolution du taux de variation a aussi un impact

sur les estimations avec notamment une structure d'influences temporelles qui peut être biaisée.

TABLE 4.5 – Résultats des simulations en cas de mauvaise spécification des trajectoires des processus (scénario 6) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|---------------|----------------|--------------------|---------------|--------------|-----------------------|--------------------|
| γ_{01} | 0.005 | -0.035 | 782.3 | 0.008 | 0.007 | 1.0 |
| γ_{02} | -0.076 | -0.138 | 82.4 | 0.015 | 0.014 | 1.0 |
| $L.u_1v_1$ | 0.221 | 0.177 | 20.1 | 0.034 | 0.036 | 75.0 |
| $L.u_2v_2$ | 0.181 | 0.108 | 40.0 | 0.025 | 0.026 | 16.0 |
| $l.u_1$ | 0.128 | 0.115 | 10.3 | 0.012 | 0.012 | 75.0 |
| $L.u_2$ | 0.215 | 0.177 | 18.0 | 0.015 | 0.016 | 29.0 |
| a_{11} | -0.199 | -0.157 | 21.0 | 0.034 | 0.037 | 74.0 |
| a_{12} | 0.086 | 0.086 | 0.1 | 0.011 | 0.012 | 89.0 |
| a_{21} | 0.184 | 0.156 | 14.9 | 0.015 | 0.014 | 54.0 |
| a_{22} | -0.219 | -0.134 | 39.0 | 0.027 | 0.029 | 12.0 |
| σ_1 | 0.305 | 0.307 | 0.6 | 0.011 | 0.011 | 96.0 |
| σ_2 | 0.524 | 0.529 | 1.0 | 0.021 | 0.021 | 95.0 |
| η_{01} | 0.992 | 1.127 | 13.5 | 0.150 | 0.153 | 82.0 |
| η_{11} | 0.871 | 0.975 | 11.9 | 0.092 | 0.088 | 81.0 |
| η_{02} | 1.817 | 1.819 | 0.1 | 0.030 | 0.029 | 97.0 |
| η_{12} | 1.389 | 1.385 | 0.2 | 0.025 | 0.026 | 96.0 |

Scénario 7 : Mauvaise spécification des corrélations initiales entre les processus

Ici nous explorons l'effet de la non prise en compte des corrélations entre les niveaux initiaux des processus latents. Pour cela, nous reprenons le modèle du scénario 1 de la section 4.0.2. Les données sont simulées de sorte que les effets aléatoires u_{i1}, u_{i2}, u_{i3} soient corrélés. Ensuite nous estimons sur les données simulées un modèle dans lequel les effets aléatoires u_{i1}, u_{i2}, u_{i3} sont indépendants. Les résultats de ce cas d'étude sont résumés dans le tableau 4.6. La non prise en compte des corrélations entre les niveaux initiaux

SIMULATIONS

des processus entraîne des biais sur les paramètres d'évolution des trajectoires et sur les paramètres d'influences temporelles entre les processus comme attendu.

TABLE 4.6 – Résultats des simulations en cas d'omission des corrélations entre les niveaux initiaux des processus latents (scénario 7) : Estimation moyenne, biais, écart-type (ET) empirique, écart-type (ET) asymptotique moyen et taux de couverture de 100 simulations mimant un échantillon de 512 sujets du jeu de données ADNI 1 ; estimation des paramètres du modèles sans la prise en compte des corrélations résiduelles au temps $t = 0$ entre les niveaux des processus latents.

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|---------------|-------------------|-----------------------|------------------|-----------------|--------------------------|-----------------------|
| β_{11} | -0.627 | -0.593 | 5.5 | 0.152 | 0.071 | 92.5 |
| β_{21} | -0.841 | -0.836 | 0.6 | 0.101 | 0.097 | 96.8 |
| β_{12} | -0.317 | -0.192 | 39.5 | 0.042 | 0.047 | 26.9 |
| β_{22} | -1.801 | -1.711 | 5.0 | 0.112 | 0.114 | 89.2 |
| β_{13} | 0.001 | 0.012 | 1115.0 | 0.070 | 0.080 | 97.8 |
| β_{23} | -1.886 | -1.801 | 4.5 | 0.121 | 0.130 | 90.3 |
| γ_{01} | -0.084 | -0.086 | 2.0 | 0.014 | 0.015 | 97.8 |
| γ_{11} | -0.037 | -0.028 | 25.1 | 0.035 | 0.022 | 88.2 |
| γ_{21} | 0.007 | 0.008 | 8.1 | 0.004 | 0.004 | 92.5 |
| γ_{31} | 0.004 | 0.004 | 4.3 | 0.005 | 0.005 | 94.6 |
| γ_{02} | 0.119 | 0.119 | 0.0 | 0.026 | 0.030 | 96.8 |
| γ_{12} | -0.443 | -0.499 | 12.6 | 0.063 | 0.068 | 87.1 |
| γ_{22} | -0.014 | -0.018 | 26.5 | 0.006 | 0.006 | 91.4 |
| γ_{32} | 0.001 | -0.002 | 341.0 | 0.009 | 0.008 | 89.2 |
| γ_{03} | 0.027 | 0.028 | 2.9 | 0.050 | 0.048 | 95.7 |
| γ_{13} | 0.171 | -0.221 | 229.1 | 0.071 | 0.070 | 0.0 |
| γ_{23} | 0.021 | -0.021 | 199.8 | 0.012 | 0.011 | 4.3 |
| γ_{33} | 0.005 | -0.012 | 340.4 | 0.015 | 0.015 | 78.5 |
| $L.v1u1$ | 0.340 | 0.053 | 84.5 | 0.030 | 0.014 | 0.0 |
| $L.v2u2$ | 0.301 | 0.335 | 11.2 | 0.044 | 0.039 | 87.1 |
| $L.v3u3$ | 0.122 | 0.144 | 18.0 | 0.035 | 0.035 | 92.5 |
| $L.v1$ | 0.058 | 0.057 | 1.5 | 0.008 | 0.004 | 88.2 |
| $L.v2$ | 0.177 | 0.175 | 1.2 | 0.018 | 0.019 | 95.7 |
| $L.v3$ | 0.235 | 0.240 | 2.2 | 0.025 | 0.025 | 95.7 |

Suite de la TABLE 4.6- Résultats des simulations en cas d'omission des corrélations entre les niveaux initiaux des processus latents (scénario 7).

| paramètre | valeur simulée | estimation moyenne | biais relatif | ET empirique | ET asymptotique moyen | taux de couverture |
|-------------|-------------------|-----------------------|------------------|-----------------|--------------------------|-----------------------|
| a_{11}^0 | -0.045 | 0.053 | 218.5 | 0.029 | 0.014 | 0.0 |
| a_{12}^0 | 0.003 | 0.006 | 103.5 | 0.005 | 0.005 | 89.2 |
| a_{12}^1 | 0.025 | 0.026 | 2.9 | 0.008 | 0.007 | 89.2 |
| a_{13}^0 | 0.005 | 0.006 | 12.8 | 0.006 | 0.006 | 91.4 |
| a_{13}^1 | -0.004 | -0.004 | 9.4 | 0.006 | 0.007 | 97.8 |
| a_{21}^0 | 0.148 | 0.058 | 61.0 | 0.017 | 0.018 | 0.0 |
| a_{21}^1 | 0.024 | 0.028 | 15.6 | 0.024 | 0.021 | 90.3 |
| a_{22}^0 | -0.301 | -0.320 | 6.4 | 0.043 | 0.039 | 91.4 |
| a_{23}^0 | 0.064 | 0.074 | 15.8 | 0.018 | 0.017 | 90.3 |
| a_{23}^1 | 0.011 | 0.011 | 3.4 | 0.019 | 0.018 | 91.4 |
| a_{31}^0 | 0.053 | 0.046 | 13.2 | 0.021 | 0.023 | 96.8 |
| a_{31}^1 | 0.041 | 0.047 | 14.4 | 0.029 | 0.029 | 92.5 |
| a_{32}^0 | 0.049 | 0.042 | 15.2 | 0.020 | 0.022 | 95.7 |
| a_{32}^1 | 0.086 | 0.088 | 2.0 | 0.026 | 0.027 | 96.8 |
| a_{33}^0 | -0.136 | -0.167 | 22.9 | 0.038 | 0.037 | 83.9 |
| σ_1 | 0.235 | 0.161 | 31.4 | 0.172 | 0.008 | 77.4 |
| σ_2 | 0.404 | 0.276 | 31.6 | 0.264 | 0.014 | 53.8 |
| σ_3 | 0.732 | 0.310 | 57.7 | 0.624 | 0.031 | 50.5 |
| η_{01} | 3.193 | 3.173 | 0.6 | 0.289 | 0.310 | 96.8 |
| η_{11} | 4.197 | 4.189 | 0.2 | 0.158 | 0.135 | 89.2 |
| η_{02} | 3.821 | 3.836 | 0.4 | 0.201 | 0.200 | 94.6 |
| η_{12} | 2.466 | 2.608 | 5.8 | 0.084 | 0.090 | 60.2 |
| η_{03} | 2.501 | 2.497 | 0.1 | 0.117 | 0.124 | 97.8 |
| η_{13} | 1.376 | 1.446 | 5.1 | 0.055 | 0.058 | 77.4 |

Conclusion

Cette deuxième partie de l'étude de simulation avait pour but d'évaluer les problèmes d'inférence possibles en présence d'une mauvaise spécification du modèle. D'après les résultats de ces simulations, il faut être vigilant quant à la spécification du modèle. En effet, l'omission d'une variable explicative importante ou d'une trajectoire plus complexe

peut avoir des conséquences sur la structure d'influences temporelles estimée.

4.0.4 Mauvaise spécification de modèle et impact de la discrétisation sur la structure d'influences temporelles entre les processus latents

Il est légitime de se poser la question à savoir si une structure d'influences temporelles entre des processus latents inférée dans notre approche de modélisation en temps discret reflète la même structure d'influences temporelles en temps continu qui aurait pu être obtenue par un modèle à équations différentielles et qui aurait pu bénéficier d'une interprétation causale. Pour cela, nous avons évalué l'erreur de première espèce associée à chaque paramètre d'influence temporelle dans notre approche de modélisation pour plusieurs discrétisations du temps lorsque les données (et la structure d'influences temporelles) ont été générées en temps continu.

Nous avons considéré un réseau de trois processus latents définis en temps continu et pouvant s'influencer mutuellement à travers une matrice d'influences $A = (a_{dd'})_{(d,d')=\{1,2,3\}^2}$ mesurant l'influence du processus d' sur le processus d . Nous avons considéré 6 scénarios. Dans chaque scénario, une seule influence temporelle entre deux processus latents d et d' est inexistante à l'échelle infinitésimale (en temps continu) $a_{dd'} = 0$. Nous estimons ensuite le modèle sur les données générées en temps continu avec 3 discrétisations du temps $\delta = 1/3, 1/2, 1$.

Pour chaque scénario et chaque pas de discrétisation, l'erreur de première espèce associée au paramètre d'influence temporelle inexistante est calculée. Cette erreur évalue la probabilité que l'on a de rejeter l'inexistence (pourtant réelle) du lien d'influence $a_{dd'}$ lorsque l'on discrétise le temps. Nous considérons une valeur nominale de 5%. Sur $R=200$ répliques, nous nous attendons donc à avoir une erreur de première espèce dans l'intervalle de confiance à 95% de l'erreur nominale attendue, soit : [1.98% , 8.02%].

Les données de cette étude de simulation sont simulées en temps quasi continu à l'aide d'un solveur [Soetaert et al., 2010] avec un pas $\delta = 0.001$. Les paramètres de génération sont définis à partir des paramètres estimés en temps discret ($\delta = 1$) sur l'échantillon de jeu de données ADNI 1. Ensuite les paramètres de simulation pour le pas $\delta = 0.001$ (approximation du temps continu) ont été déduits à l'aide des relations établies entre les paramètres estimés d'un même modèle pour deux pas différents de discrétisation (section 3.3 du chapitre 3).

Pour générer les données nous avons considéré un modèle non ajusté sur des variables explicatives et un taux de variation constant des processus latents entre deux temps t et $t + 0.001$. La structure d'influences temporelles entre les processus est aussi supposée constante. Le modèle de génération des données pour $t_{ij} > 0$ et $\delta = 0.001$ est décrit comme suit :

$$\begin{cases} \Lambda_{i1}(0) = \beta_1 + u_{i1} \\ \Lambda_{i2}(0) = \beta_2 + u_{i2} \\ \Lambda_{i3}(0) = \beta_3 + u_{i3} \\ \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_1 + v_{i1} + a_{11}\Lambda_{i1}(t_{ij}) + a_{12}\Lambda_{i2}(t_{ij} - \delta) + a_{13}\Lambda_{i3}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_2 + v_{i2} + a_{21}\Lambda_{i1}(t_{ij} - \delta) + a_{22}\Lambda_{i2}(t_{ij} - \delta) + a_{23}\Lambda_{i3}(t_{ij} - \delta) \\ \frac{\Delta\Lambda_{i3}(t_{ij})}{\delta} = \gamma_3 + v_{i3} + a_{31}\Lambda_{i1}(t_{ij} - \delta) + a_{32}\Lambda_{i2}(t_{ij} - \delta) + a_{33}\Lambda_{i3}(t_{ij} - \delta) \end{cases} \quad (4.5)$$

$u_{i1} \perp\!\!\!\perp u_{i2} \perp\!\!\!\perp u_{i3}$
 $\frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} = \Lambda_{ik}(t_{ijk}) + \tilde{\epsilon}_{ijk}, \quad k = 1, 2, 3$

Les résultats sont présentés dans le tableau 4.7. Quel que soit le niveau de discrétisation, l'inférence réalisée en temps discret semble correcte. L'approche discrétisée met correctement en évidence l'existence d'influences temporelles définies au niveau infinitésimal.

TABLE 4.7 – Impact de la discrétisation sur la structure d'influences temporelles entre les processus : Erreur de type I (en %) associée à chaque paramètre d'influence $a_{dd'}$ avec $(d, d') = \{1, 2, 3\}^2$ et $d \neq d'$ mesurant l'influence du processus d' sur le procesus d pour des modèles estimés avec les pas de discrétisation $\delta = 1/3, 1/2, 1$ sur un échantillon simulé en temps quasi continu avec un solveur ($\delta = 0.001$). L'intervalle de confiance à 95% de l'erreur de type -I attendue est de $[1.98, 8.02]$ pour l'erreur nominale de 5%.

| Parametre | $\delta = 1/3$ | $\delta = 1/2$ | $\delta = 1$ |
|-----------|----------------|----------------|--------------|
| a_{12} | 5.5 | 4.5 | 6.5 |
| a_{13} | 3.5 | 5.5 | 5.0 |
| a_{21} | 7.0 | 7.5 | 8.0 |
| a_{23} | 5.0 | 4.5 | 5.0 |
| a_{31} | 4.0 | 3.5 | 5.5 |
| a_{32} | 7.0 | 7.5 | 5.5 |

Conclusion

SIMULATIONS

Le modèle dynamique défini en temps discret permet de retrouver correctement des structures d'influences temporelles telles que définies au niveau infinitésimal dans des modèles à équations différentielles.

5 Applications

Ce chapitre est consacré à l'application de la méthodologie développée dans la thèse. Nous présentons deux applications réalisées sur les données de la cohorte ADNI. L'objectif principal de ces applications est d'explorer les structures d'influences temporelles entre les dimensions anatomique, cognitive et fonctionnelle ainsi que leur évolution au cours de la maladie d'Alzheimer.

Dans la première application décrite aussi dans le papier méthodologique issu de la thèse (voir annexe), les dimensions anatomique, cognitive et fonctionnelle sont définies de façon assez générale. Cette application a servi de base pour la génération des données dans les études de simulation et nous a permis d'illustrer la méthodologie. La deuxième application vise à affiner les résultats obtenus en révisant les dimensions latentes considérées pour explorer plus finement les relations entre structure du cerveau, fonctions cognitives et autonomie.

5.1 Données d'application : ADNI 1

Dans les deux applications, nous avons utilisé les données de la phase ADNI 1 de l'initiative ADNI.

5.1.0.1 Initiative internationale ADNI

Comme évoqué en introduction, ADNI est une étude longitudinale multi-sites dont l'objectif est de collecter sur le long terme des marqueurs génétiques, cliniques, biochimiques et d'imagerie (IRM) pour mieux comprendre la progression des atteintes clinico-pathologiques dans la maladie d'Alzheimer.

Trois groupes de sujets sont inclus dans l'étude ADNI sur la base de leur stade clinique de la maladie d'Alzheimer à la date d'inclusion. Sont principalement utilisés pour l'inclusion le Mini Mental State Examination (MMSE), un test neuropsychologique sur 30 points qui évalue le fonctionnement cognitif global [Folstein et al., 1975], et la Clinical Dementia Rating (CDR), une échelle sur 5 points qui évalue la sévérité des symptômes de démence [Morris, 1993].

APPLICATIONS

- Le groupe CN (Cognitively Normal) regroupe les sujets cognitivement normaux qui ne présentent pas de déficits cognitifs. Ils ont un score entre 24 et 30 au MMSE, un score de 0 à la CDR et ne présentent pas de signe de dépression.
- Le groupe MCI (Mild Cognitive Impairment) regroupe les sujets présentant des plaintes cognitives et de légers déficits de mémoire objectivés et ne présentant pas de déficits significatifs d'autres fonctions cognitives, sans répercussion sur les activités de la vie quotidienne et qui ne sont pas diagnostiqués déments. Ils ont un score au MMSE entre 24 et 30, un score à la CDR de 0.5.
- Le groupe AD (Alzheimer's dementia) regroupe les sujets diagnostiqués au stade démentiel de la maladie d'Alzheimer. Ils ont un score au MMSE entre 20 et 26, un score à la CDR de 0.5 ou 1 et satisfont le critère NINCDS/ADRDA de diagnostic de la démence de type Alzheimer.

L'initiative ADNI est un projet évolutif avec plusieurs phases. Elle est actuellement à la phase ADNI2 après complétion des phases ADNI 1 et ADNI GO. À chaque phase, des sujets supplémentaires sont recrutés.

Pour ce chapitre, nous avons considéré uniquement les données de la phase ADNI 1 qui inclut environ 800 individus âgés de 55 à 90 ans dont 200 sujets sont au stade CN, 400 sont au stade MCI et 200 sont au stade AD. À l'inclusion, les informations socio-démographiques, des bio-marqueurs biochimiques et d'imagerie médicale, des tests cognitifs et fonctionnels sont collectés. Les sujets inclus sont ensuite suivis pendant trois ans pour les groupes CN et MCI et pendant deux ans pour le groupe AD. Le protocole prévoit une visite tous les six mois. À chaque visite, des bio-marqueurs biochimiques et d'imagerie médicale, des tests cognitifs et fonctionnels sont collectés ainsi que les données sur la progression des sujets entre les trois stades de la MA Mueller et al. [2005].

5.1.0.2 Échantillon d'étude

La sélection de l'échantillon d'étude a été faite sur la base des variables explicatives et des marqueurs d'intérêt. L'échantillon inclut tous les sujets de la cohorte ADNI 1 :

- n'ayant pas de données manquantes aux variables explicatives sexe (**gender**), âge à l'inclusion (**apoe4**), le nombre d'années d'étude (**educ12**, variable binaire qui vaut 1 si le nombre d'années d'étude est supérieur à 12 ans), l'allèle apolipoprotéine $\epsilon 4$ (**apoe4**, variable binaire indiquant si l'individu est porteur d'au moins une copie de l'allèle apolipoprotéine $\epsilon 4$), et les variables binaires **MCI** et **AD** qui codent le stade clinique de la maladie ;

- ayant au moins une mesure aux marqueurs des trois dimensions d'intérêt.

Les marqueurs considérés ici sont les marqueurs des dimensions anatomique, cognitive et fonctionnelle disponibles au moment de l'acquisition de la base de données ADNI 1.

Pour la dimension cognitive, les dix-neuf scores cognitifs réalisés dans ADNI 1 ont été considérés (suivant [Crane et al., 2012]) à savoir :

- le test de fluence verbale (catégories animale et végétale),
- cinq sous-composantes du test ADAS (pour « Alzheimer Disease Assessment Scale »),
- cinq sous-composantes du test AVLT (pour « Auditory Verbal Learning Test »)
- des tests des fonctions exécutives dont le test de la montre, le « Trail Making Test (A et B) », le « Digit Span Test » (« Forward » et « Backward ») et enfin le « Digit Symbol Substitution Test ».

Pour la dimension fonctionnelle, la FAQ (pour « Functional Assessment Questionnaire ») a été collectée; elle mesure l'autonomie dans les activités instrumentales de la vie quotidienne (IADL) (telles que la préparation de repas, la gestion des finances, etc.) à travers 30 items [Pfeffer et al., 1982].

Les examens d'imagerie médicale (IRM) ont donné lieu à un nombre extrêmement important de mesures notamment de la structure anatomique du cerveau. Pour appréhender la dimension anatomique, nous nous sommes focalisés sur les données répétées collectées par IRM (Freesurfer version 4.4.0) des volumes relatifs de 9 régions corticales précédemment identifiées comme pertinentes dans la maladie d'Alzheimer (temporale pôle, temporale médiane, temporale inférieure, marginale supérieure, frontale supérieure, frontale inférieure, pariétale supérieure, pariétale inférieure, précuneus) [Dickerson et al., 2008] ainsi qu'au volume relatif de l'hippocampe dans les hémisphères droit et gauche, l'hippocampe étant la zone du cerveau qui est fortement associée à la mémoire épisodique [Grady et al., 2003].

Dans la première application sur les données de ADNI 1, nous avons construit un composite score pour chaque dimension. Le composite score est la moyenne des scores standardisés des marqueurs de la dimension sauf pour la dimension fonctionnelle pour laquelle nous avons utilisé comme score la somme standardisée des 30 items. Ces marqueurs sont nommés respectivement **Anat10**, **Cog19**, **Fct**.

Dans la deuxième application, plutôt que de considérer des scores composites, nous avons directement considéré des marqueurs spécifiques à savoir :

- le score **Hip** comme moyenne des volumes relatifs standardisés des hippocampes

APPLICATIONS

- droit et gauche pour mesurer la dimension « atrophie de l'hippocampe »,
- le score composite **Avlt** comme moyenne des sous-scores standardisées du « rappel immédiat », « du rappel différé », et « de la reconnaissance différée de mots » du test de mémoire épisodique « Auditory Verbal Learning Test »,
 - le score standardisé **CaFlu** au test de fluence verbale (catégorie animale)

5.1.0.3 Description de l'échantillon

Au total 656 sujets ont été inclus dans l'échantillon dont 190 sujets contrôles (CN), 332 sujets au stade MCI et 144 sujets au stade AD. Les caractéristiques socio-démographiques des individus de l'échantillon à l'inclusion, ainsi que le suivi des individus et les marqueurs répétés d'intérêt sont décrits dans le tableau 5.1. La moyenne d'âge à l'inclusion des sujets est entre 75 et 76 ans dans les trois groupes. Le groupe MCI inclut plus d'hommes que les deux autres groupes (64 % versus 51.1 % et 50.7 % chez les CN et AD). On observe des gradients selon le niveau d'études et la présence de l'allèle apolipoprotéine $\epsilon 4$ avec 91.1%, 82.3% et 72.2% des sujets ayant fait plus de 12 ans d'étude chez les CN, MCI et AD respectivement, et 24.2 %, 56.2 % et 67.4% de porteurs de l'allèle apolipoprotéine $\epsilon 4$.

TABLE 5.1 – Statistiques descriptives de l'échantillon de 656 sujets de la base de données ADNI 1 répartis en trois groupes de sujets : CN, MCI et AD.

| | | Ech. CN (n = 190) | Ech. MCI (n = 322) | Ech. AD (n = 144) |
|---|--|--------------------------|--------------------|----------------------|
| | | stat. | stat. | stat. |
| carac. socio-démographiques | | | | |
| age en année : \bar{x}^a (min max) | | 76.1(62 89.6) | 75.0(55.0 88.8) | 75.7(56.4 90.9) |
| Homme : eff. ^b (%) | | 97(51.1) | 206(64.0) | 73(50.7) |
| années d'étude >12 ans | | 173(91.1) | 265(82.3) | 104(72.2) |
| porteur apo $\epsilon 4$ | | 46(24.2) | 181(56.2) | 97(67.4) |
| inclusion : tx^c = 1.72% | | corr.^d | n = 190 | n = 322 |
| score Anat10 : \bar{x} (min max) | | | 0.48(-2.02 1.67) | -0.02(-2.86 1.63) |
| score Cog19 : \bar{x} (min max) | | 0.5 | 0.58(-0.41 1.35) | -0.05(-1.36 1.42) |
| score Fct : \bar{x} (min max) | | 0.4 0.6 | 0.72(-0.20 0.74) | 0.13(-2.58 0.74) |
| score Hip : \bar{x} (min max) | | | 0.74(-1.27 3.41) | -0.14(-21.92 3.09) |
| score Avlt : \bar{x} (min max) | | | 0.87(-0.77 2.33) | -0.18(-1.54 2.33) |
| score CatFlu : \bar{x} (min max) | | | 0.60(-1.69 3.77) | -0.06(-2.04 2.36) |
| Suivi à 6 mois, tx = 3.25% | | | n = 190 | n = 321 |
| score Anat10 : \bar{x} (min max) | | | 0.42(-2.56 1.59) | -0.125(-2.79 1.57) |
| score Cog19 : \bar{x} (min max) | | 0.6 | 0.58(-0.45 1.63) | -0.13(-1.78 1.52) |
| score Fct : \bar{x} (min max) | | 0.5 0.6 | 29.7(24 30) | 25.2(7 30) |
| score Hip : \bar{x} (min max) | | | 0.68(-1.56 3.47) | -0.22(-2.03 3.00) |
| score Avlt : \bar{x} (min max) | | | 0.72(-1.24 2.27) | -0.32(-1.80 2.27) |
| score CatFlu : \bar{x} (min max) | | | 0.65(-1.87 3.95) | -0.13(-2.22 2.71) |

Suite de la Table 5.1- Statistiques descriptives de l'échantillon de 656 sujets de la base de données ADNI 1 répartis en trois groupes de sujets : CN, MCI et AD.

| | | Ech. CN (n = 190) | Ech. MCI (n = 322) | Ech. AD (n = 144) |
|--|---------|---------------------|----------------------|----------------------|
| | | stat. | stat. | stat. |
| Suivi à 12 mois, tx = 4.61% | | n = 189 | n = 319 | n = 143 |
| score Anat10 : \bar{x} (min max) | | 0.38(-2.65 1.70) | -0.22(-2.61 1.60) | -0.84(-2.40 1.47) |
| score Cog19 : \bar{x} (min max) | 0.6 | 0.64(-0.72 1.69) | -0.12(-1.69 1.53) | -0.76(-2.41 0.42) |
| score Fct : \bar{x} (min max) | 0.5 0.7 | 0.71(-0.83 0.744) | -0.16(-3.37 .74) | -1.93(-4 0.586) |
| score Hip : \bar{x} (min max) | | 0.65(1.76 3.67) | -0.31(-2.21 3.11) | -0.89(-2.68 3.78) |
| score Avlt : \bar{x} (min max) | | 0.90(-0.81 2.54) | -0.28(-2.03 2.42) | -1.01(-2.15 0.74) |
| score CatFlu : \bar{x} (min max) | | 0.69(-1.87 3.95) | -0.18(-2.39 7.64) | -0.91(-2.57 1.30) |
| Suivi à 18 mois, tx = 8.19% | | n = 0 | n = 294 | n = 7 |
| score Anat10 : \bar{x} (min max) | | -(- -) | -0.25(-2.38 1.56) | 0.05(0.05 0.05) |
| score Cog19 : \bar{x} (min max) | 0.5 | -(- -) | -0.18(-1.97 1.62) | -0.27(-0.59 0.02) |
| score Fct : \bar{x} (min max) | 0.3 0.5 | -(- -) | -0.31(-3.68 0.744) | -1.59(-3.68 0.11) |
| score Hip : \bar{x} (min max) | | -(- -) | -0.35(-2.14 3.15) | -0.42(-0.42 -0.42) |
| score Avlt : \bar{x} (min max) | | -(- -) | -0.39(-2.03 2.45) | -0.92(-2.06 0.23) |
| score CatFlu : \bar{x} (min max) | | -(- -) | -0.27(-2.57 3.59) | -0.96(-2.92 -0.10) |
| Suivi à 24 mois, tx = 10.97% | | n = 179 | n = 274 | n = 127 |
| score Anat10 : \bar{x} (min max) | | 0.30(-2.08 1.57) | -0.30(-2.57 1.51) | -1.01(-2.52 0.98) |
| score Cog19 : \bar{x} (min max) | 0.6 | 0.65(-0.54 1.80) | -0.13(-1.99 1.64) | -0.96(-2.52 0.18) |
| score Fct : \bar{x} (min max) | 0.6 0.8 | 0.68(-0.99 0.74) | -0.53(-4.00 0.74) | -2.41(-4.00 0.59) |
| score Hip : \bar{x} (min max) | | 0.54(-1.70 3.39) | -0.41(-2.18 2.91) | -1.05(-2.75 3.77) |
| score Avlt : \bar{x} (min max) | | -0.98(-1.92 2.48) | -0.38(-2.12 2.51) | -1.16(-2.15 0.03) |
| score CatFlu : \bar{x} (min max) | | 0.71(-1.34 3.77) | -0.30(-2.57 3.59) | -1.25(-2.92 1.48) |
| Suivi à 36 mois, tx = 18.41% | | n = 165 | n = 224 | n = 11 |
| score Anat10 : \bar{x} (min max) | | 0.32(-1.53 2.31) | -0.35(-2.76 1.50) | 0.32(-0.36 1.00) |
| score Cog19 : \bar{x} (min max) | 0.5 | 0.65(-0.47 1.75) | -0.21(-2.45 1.66) | -1.10(-1.55 -0.74) |
| score Fct : \bar{x} (min max) | 0.5 0.7 | 0.61(-2.58 0.74) | -0.85(-4.00 0.74) | -2.63(-4.00 0.428) |
| score Hip : \bar{x} (min max) | | 0.55(-1.22 4.29) | -0.53(-2.34 2.64) | -1.17(-1.53 -0.82) |
| score Avlt : \bar{x} (min max) | | -0.79(-1.13 2.34) | -0.49(-2.09 2.39) | -1.75(-2.03 -1.49) |
| score CatFlu : \bar{x} (min max) | | 0.69(-1.34 4.30) | -0.46(-2.92 3.24) | -1.55(-2.57 0.81) |
| Coefficient de variation | | | | |
| score Anat10 : cv, et. ^e | | 2.11, 0.81 | -4.69, 0.90 | -1.18, 0.90 |
| score Cog19 : cv, et. | | 0.97, 0.60 | -5.59, 0.72 | -0.97, 72 |
| score Fct : cv, et. | | 0.71, 0.49 | 4.11, 1.03 | 0.61, 1.10 |
| score Hip : cv, et. | | 1.42, 0.91 | -3.13, 0.96 | -1.14, 0.93 |
| score Avlt : cv, et. | | 0.96, 0.82 | -2.71, 0.88 | -0.76, 0.74 |
| score CatFlu : cv, et. | | 1.48, 0.99 | -4.63, 0.98 | -1.05, 0.94 |
| Nombre de mesures | | | | |
| score Anat10 : med. ^f (min max) | | 5(1 5) | 5(2 6) | 4(1 5) |
| score Cog19 : med.(min max) | | 5(1 5) | 5(2 6) | 3(0 6) |
| score Fct : med.(min max) | | 5(2 5) | 6(3 6) | 4(3 6) |
| score Hip : med.(min max) | | 5(2 5) | 5(2 6) | 4(1 6) |
| score Avlt : med.(min max) | | 5(3 5) | 5(2 6) | 4(3 6) |
| score CatFlu : med.(min max) | | 5(1 5) | 6(3 6) | 4(1 6) |

APPLICATIONS

Suite de la Table 5.1- Statistiques descriptives de l'échantillon de 656 sujets de la base de données ADNI 1 répartis en trois groupes de sujets : CN, MCI et AD.

| | Ech. CN (n = 190) | Ech. MCI (n = 322) | Ech. AD (n = 144) |
|-------------------------------------|-------------------|--------------------|-------------------|
| | stat. | stat. | stat. |
| Nombre visite : med. (q1 q3) | 5(5 5) | 6(5 6) | 4(4 4) |

^a \bar{x} : moyenne, ^b eff. : effectif, ^c taux de données manquantes à chaque visite concernée,

^d du haut vers le bas et de gauche vers la droite, les corrélations respectivement entre le marqueur anatomique et le marqueur cognitif, entre le marqueur anatomique et le marqueur fonctionnel, entre le marqueur cognitif et le marqueur fonctionnel. ^e cv : coefficient de variation, et. : écart-type, med.^f : médiane.

Nous avons calculé le taux de visites manquantes par rapport au nombre maximal de visite de 7 (une visite à l'inclusion et une visite tous les 6 mois sur 3 ans de suivi). Le taux de visite manquante pour les sujets de l'étude est de 18% en moyenne. Ce taux relativement élevé est dû au protocole de l'étude ADNI 1 qui prévoit des visites à l'inclusion pour tous les sujets puis à 6, 12, 24 et 36 mois pour les sujets normaux, à 6, 12, 18, 24, et 36 mois pour les MCI et à 6, 12, et 24 mois pour les sujets AD. Toutefois 7 sujets AD ont des visites à 18 mois et 11 sujets AD à 36 mois. Nous avons tenu compte de ces visites additionnelles dans les analyses.

Le nombre moyen de visite est respectivement de 5, 6 et 4 dans les groupes CN, MCI et AD. Le nombre médian de mesures varie entre 4 à 5 pour la dimension anatomique, entre 3 et 5 pour la dimension cognitive et entre 4 et 6 pour le score fonctionnel.

La distribution tout au long du suivi de chaque score est donnée en Figure 5.1 (colonne de gauche). Seul le score fonctionnel montre une asymétrie très importante et un écart flagrant à la distribution normale. Nous avons cependant systématiquement considéré les fonctions de lien approchées par I-splines pour les trois dimensions.

Notons (tableau 5.1) que les variations des scores composites sont plus importantes chez les sujets MCI que chez les CN et AD tout au long du suivi, et que, relativement à leur moyenne, les scores composites cognitif et anatomique varient beaucoup plus que le score fonctionnel.

Les trois marqueurs composites utilisés pour la première application sont globalement très corrélés entre eux à l'inclusion dans l'étude avec des corrélations entre 0.3 et 0.6. Le coefficient de corrélation est plus important entre le marqueur cognitif et les marqueurs anatomique et fonctionnel, ces deux derniers étant moins corrélés.

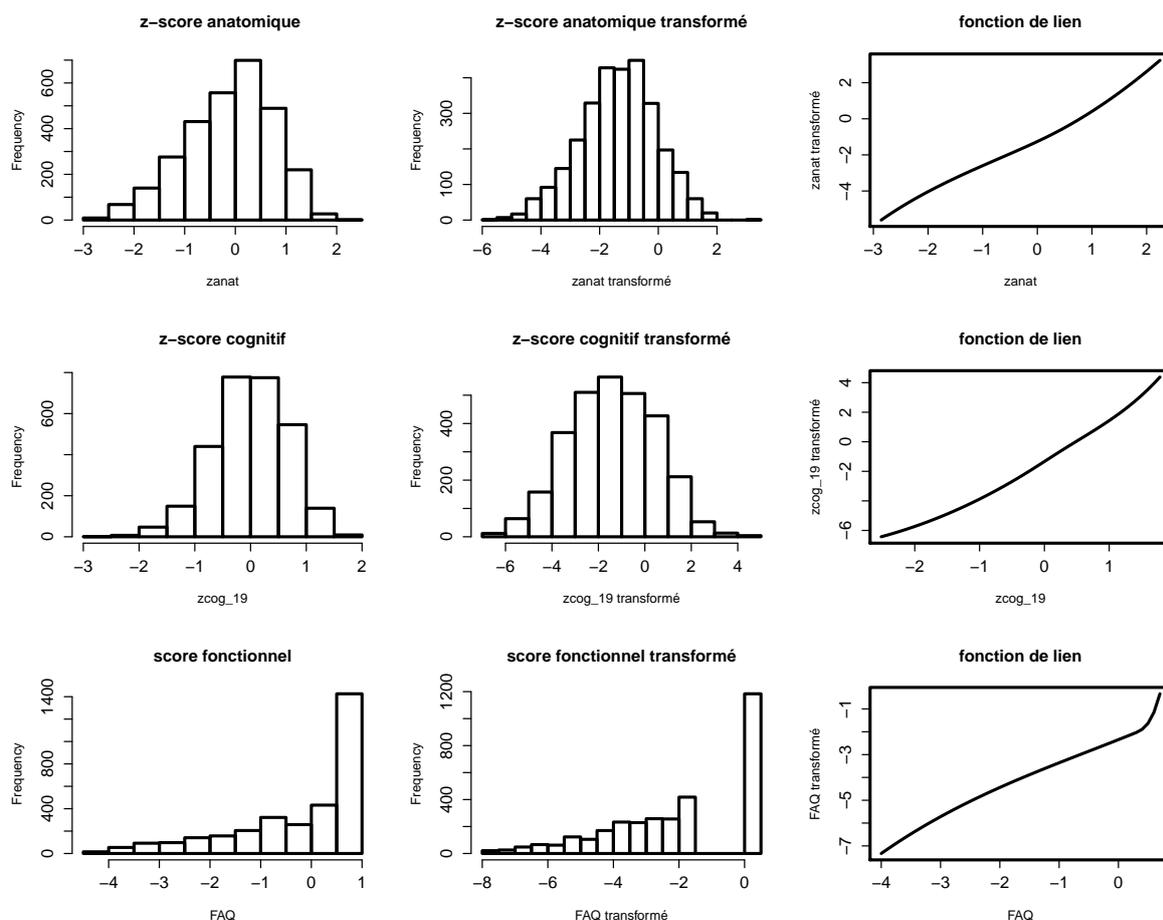


FIGURE 5.1 – Histogrammes des marqueurs dans les échelles naturelle et transformée, et fonctions de lien estimées pour les dimensions anatomique, cognitive et fonctionnelle).

5.2 Application aux dimensions anatomique, cognitive et fonctionnelle globales

Dans cette première application, nous avons considéré les composites scores décrits plus haut pour résumer les dimensions anatomique, cognitive et fonctionnelle.

5.2.1 Modélisation

5.2.1.1 Modèle

Pour cette application, nous avons considéré les spécifications suivantes :

- le temps est en années; nous avons considéré le temps exact de mesure depuis l'inclusion et non le temps en terme de visites théoriques;

APPLICATIONS

- le pas de discrétisation $\delta = 0.23$; cela correspond à environ 3 mois. Nous n'avons pas considéré un pas plus grand pour éviter d'avoir deux mesures dans le même intervalle de temps.
- le niveau initial des dimensions est ajusté sur les variables explicatives AGE0, gender, educ12, apoe4, MCI et AD ;
- le taux de variation entre deux temps consécutif t et $t + \delta$ a une évolution linéaire selon le temps et est ajusté sur les variables explicatives gender, educ12, apoe4, MCI et AD ;
- la structure d'influences temporelles entre les dimensions est ajustée sur les variables MCI et AD.
- un intercept aléatoire sur le niveau initial de chaque dimension et deux effets aléatoires sur le taux de variation (intercept et temps) sont spécifiés par dimension.

Nous avons choisi des fonctions de lien par I-splines intégrés pour relier chaque marqueur composite à la dimension mesurée. Comme mentionné en méthodologie, une sélection des variables explicatives a été réalisée dans des analyses préliminaires univariées (séparées pour chaque dimension).

Le modèle structurel final est défini comme suit :

$$\left\{ \begin{array}{l}
 \Lambda_{i1}(0) = \beta_{01} + \beta_{11}\text{AGE0}_i + \beta_{21}\text{gender}_i + \beta_{31}\text{educ12}_i \\
 \quad + \beta_{41}\text{apoe4}_i + \beta_{51}\text{MCI}_i + \beta_{61}\text{AD}_i + u_{i1} \\
 \Lambda_{i2}(0) = \beta_{02} + \beta_{12}\text{AGE0}_i + \beta_{22}\text{gender}_i + \beta_{32}\text{educ12}_i \\
 \quad + \beta_{42}\text{apoe4}_i + \beta_{52}\text{MCI}_i + \beta_{62}\text{AD}_i + u_{i2} \\
 \Lambda_{i3}(0) = \beta_{03} + \beta_{13}\text{AGE0}_i + \beta_{23}\text{apoe4}_i + \beta_{33}\text{MCI}_i + \beta_{43}\text{AD}_i \\
 \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}\text{MCI}_i + \gamma_{21}\text{AD}_i + \gamma_{31}\text{gender}_i + \gamma_{41}\text{educ12}_i + \gamma_{51}\text{apoe4}_i + \gamma_{61}t_{ij} \\
 \quad + \gamma_{71}\text{MCI}t_{ij} + \gamma_{81}\text{AD}t_{ij} + \gamma_{91}\text{educ12}t_{ij} + \gamma_{(10)1}\text{gender}t_{ij} + v_{i11} + v_{i21}t_{ij} \\
 \quad + \alpha_{11}^0\Lambda_{i1}(t_{ij} - \delta) + (\alpha_{12}^0 + \alpha_{12}^1\text{MCI} + \alpha_{12}^2\text{AD})\Lambda_{i2}(t_{ij} - \delta) \\
 \quad + (\alpha_{13}^0 + \alpha_{13}^1\text{MCI} + \alpha_{13}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}\text{MCI}_i + \gamma_{22}\text{AD}_i + \gamma_{32}\text{gender}_i + \gamma_{42}\text{educ12}_i + \gamma_{52}\text{apoe4}_i + \gamma_{62}t_{ij} \\
 \quad + \gamma_{72}\text{MCI}t_{ij} + \gamma_{82}\text{AD}t_{ij} + \gamma_{92}\text{educ12}t_{ij} + v_{i2} + v_{i22}t_{ij} + \alpha_{22}^0\Lambda_{i2}(t_{ij} - \delta) \\
 \quad + (\alpha_{21}^0 + \alpha_{21}^1\text{MCI} + \alpha_{21}^2\text{AD})\Lambda_{i1}(t_{ij} - \delta) \\
 \quad + (\alpha_{31}^0 + \alpha_{31}^1\text{MCI} + \alpha_{31}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i3}(t_{ij})}{\delta} = \gamma_{03} + \gamma_{13}\text{MCI}_i + \gamma_{23}\text{AD}_i + \gamma_{33}\text{apoe4}_i + \gamma_{43}t_{ij} \\
 \quad + \gamma_{53}\text{MCI}t_{ij} + \gamma_{63}\text{AD}t_{ij} + \gamma_{73}\text{apoe4}t_{ij} + v_{i3} + v_{i32}t_{ij} + \alpha_{33}^0\Lambda_{i3}(t_{ij} - \delta) \\
 \quad + (\alpha_{31}^0 + \alpha_{31}^1\text{MCI} + \alpha_{31}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \quad + (\alpha_{32}^0 + \alpha_{32}^1\text{MCI} + \alpha_{32}^2\text{AD})\Lambda_{i2}(t_{ij} - \delta)
 \end{array} \right. \tag{5.1}$$

avec $\Lambda_{i1}(t_{ij}), \Lambda_{i2}(t_{ij}), \Lambda_{i3}(t_{ij})$ respectivement les dimensions anatomique, cognitive et fonctionnelle chez le sujet i au temps t_{ij} , $\Delta\Lambda_i = \Lambda_i(t_{ij}) - \Lambda_i(t_{ij} - \delta)$, et δ le pas de discrétisation ; Les vecteurs β , α et γ incluent les paramètres d'effets fixes, les paramètres β_0 sont fixés à zéro par contrainte d'identifiabilité), $(u_{i1}, u_{i2}, u_{i3}, v_{i11}, v_{i21}, v_{i21}, v_{i22}, v_{i13}, v_{i23})^\top$ vecteur d'effets aléatoires centrés de matrice de variance-covariance \mathbf{B} . \mathbf{B} suit la structure indiquée en méthodologie : les intercepts aléatoires sur le niveau initial sont standardisées, les effets aléatoires sont corrélés au sein de chaque dimension latente. Par contre, les effets aléatoires entre dimensions sont indépendants sauf les intercepts aléatoires sur le niveau initial qui captent une potentielle corrélation à l'inclusion.

5.2.1.2 Estimations des paramètres d'évolution des dimensions

Les estimations des effets fixes du modèle dynamique sont données dans le tableau 5.2. Les estimations des paramètres de la décomposition de Cholesky L de la matrice de variance-covariance des effets aléatoires sont données en annexe (tableau A 6.1). À l'inclusion, le grand âge et la présence de l'allèle apolipoprotéine $\epsilon 4$ sont associés à un niveau plus bas des dimensions anatomique et cognitive, ce qui est attendu d'après les connaissances épidémiologiques. Un nombre d'années d'étude supérieur à 12 ans est associé, comme attendu, à un niveau cognitif plus élevé mais il est aussi associé à un niveau anatomique plus faible. Ce résultat peut venir du fait que nous considérons ici principalement des épaisseurs corticales relatives au volume intra-cranien et, qu'autant le volume intra-cranien est souvent retrouvé plus grand chez les sujets de haut niveau d'étude autant il n'y a pas consensus sur des différences d'épaisseurs corticales suivant le niveau d'étude. Cela pourrait aussi expliquer pourquoi les hommes sont retrouvés avec une dimension anatomique plus faible. Toutes ces caractéristiques ne sont pas associées au niveau fonctionnel initial. Concernant les stades cliniques, on retrouve un gradient de déficit progressif dans les trois dimensions avec la sévérité des stades.

Conditionnellement à l'état actuel du système, le taux de variation de la dimension anatomique est négatif chez les sujets CN et MCI et encore plus bas chez les sujets AD (avec à noter une atténuation de la différence au cours du suivi). Cela traduit une atrophie progressive dans les trois groupes. Le taux de variation de la dimension anatomique n'est pas significativement associé au sexe, au niveau d'étude ou au temps depuis l'entrée dans l'étude (mis à part chez les AD). Seuls les porteurs de l'allèle apolipoprotéine $\epsilon 4$ ont un taux variation plus bas que les non porteurs, ce qui traduit un déclin plus important de la dimension anatomique chez les porteurs de l'allèle apolipoprotéine $\epsilon 4$ par rapport aux non-porteurs pour un même état du système.

APPLICATIONS

TABLE 5.2 – Esstimations des effets fixes des modèles du niveau initial et du taux de variation des 3 dimensions pour un pas de discrétisation $\delta = 0.23 \approx 3$ mois.

| paramètre | dim. anatomique | | | dim. cognitive | | | dim. fonctionnelle | | |
|---------------------------------|-----------------|-------|----------|----------------|-------|----------|--------------------|-------|----------|
| | θ | se | p-valeur | θ | se | p-valeur | θ | se | p-valeur |
| <u>À l'inclusion</u> | | | | | | | | | |
| MCI | -0.754 | 0.101 | < 0.001 | -1.713 | 0.124 | < 0.001 | -1.728 | 0.134 | < 0.001 |
| AD | -1.509 | 0.129 | < 0.001 | -3.182 | 0.170 | < 0.001 | -3.815 | 0.203 | < 0.001 |
| âge (en année) | -0.610 | 0.065 | < 0.001 | -0.334 | 0.055 | < 0.001 | -0.071 | 0.074 | 0.339 |
| Homme | -0.177 | 0.082 | 0.031 | -0.173 | 0.089 | 0.053 | - | - | - |
| années d'étude * | -0.255 | 0.107 | 0.018 | 0.515 | 0.118 | < 0.001 | - | - | - |
| apoe ϵ 4 | -0.293 | 0.086 | 0.001 | -0.382 | 0.096 | < 0.001 | -0.070 | 0.102 | 0.489 |
| <u>sur le taux de variation</u> | | | | | | | | | |
| Intercept | -0.133 | 0.053 | 0.013 | 0.093 | 0.118 | 0.432 | 0.104 | 0.110 | 0.342 |
| MCI | -0.052 | 0.044 | 0.234 | -0.834 | 0.197 | < 0.001 | -0.179 | 0.178 | 0.314 |
| AD | -0.201 | 0.096 | 0.036 | -1.692 | 0.396 | < 0.001 | -0.978 | 0.439 | 0.026 |
| Homme | -0.032 | 0.035 | 0.368 | 0.037 | 0.069 | 0.587 | - | - | - |
| années d'étude | -0.013 | 0.046 | 0.768 | 0.542 | 0.151 | < 0.001 | - | - | - |
| apoe ϵ 4 | -0.088 | 0.018 | < 0.001 | -0.252 | 0.081 | 0.002 | -0.285 | 0.111 | 0.010 |
| t | 0.029 | 0.029 | 0.326 | 0.008 | 0.057 | 0.884 | -0.090 | 0.053 | 0.088 |
| MCI \times t | -0.002 | 0.022 | 0.929 | -0.035 | 0.048 | 0.464 | -0.071 | 0.076 | 0.351 |
| AD \times t | 0.072 | 0.042 | 0.085 | -0.283 | 0.106 | 0.008 | 0.107 | 0.124 | 0.387 |
| Homme \times t | 0.036 | 0.020 | 0.072 | - | - | - | - | - | - |
| années d'étude \times t | -0.034 | 0.026 | 0.184 | -0.044 | 0.051 | 0.391 | - | - | - |
| apoe ϵ 4 \times t | - | - | - | - | - | - | 0.083 | 0.068 | 0.222 |

* : nombre d'années d'étude supérieur à 12 ans,

NB : Un "-" indique une variable explicative non incluse dans le modèle multivarié

après sélection des variables explicatives dans les modèles univariés.

On retrouve aussi un taux de variation de la dimension cognitive négatif chez les MCI et AD et plus bas chez les porteurs de l'allèle apolipoprotéine ϵ 4 conditionnellement à l'état actuel du système. Cependant, on trouve que les sujets avec un haut niveau d'étude ont un taux de variation de la dimension cognitive plus haut que les autres et les sujets AD ont une diminution du taux de variation de la dimension cognitive qui augmente au cours du temps par rapport aux MCI et CN.

Enfin, conditionnellement à l'état actuel du système, le taux de variation de la dimension fonctionnelle est uniquement impacté de façon significative par le stade clinique avec une chute plus grande aux stades avancés.

Pour mieux appréhender les trajectoires prédites des dimensions, la Figure 5.2 représente à chaque stade clinique de la maladie les trajectoires moyennes estimées et leur intervalles de confiances pour deux profils d'individus :

- profil 1 : femme porteuse de l'allèle apoe ϵ 4 et ayant un nombre d'années de scolarité inférieur à 12 ans

- profil 2 : femme non porteuse de l'allèle apoe ϵ 4 et ayant plus de 12 ans d'années de scolarité

À noter que sur ces figures, pour tenir compte de la variation au cours du temps très différente selon les dimensions (cf tableau 5.1), les dimensions sont ici rapportées à une échelle 0-1 (c'est-à-dire centrées sur leur valeur minimale et réduites par leur amplitude tout au long du suivi). D'une part, ces trajectoires diffèrent largement selon les deux profils de femmes, celles du profil 1 ayant systématiquement un niveau plus bas et/ou un déclin plus important que celles du profil 2. D'autre part, pour chaque dimension, les déclins s'accélèrent avec la sévérité du stade.

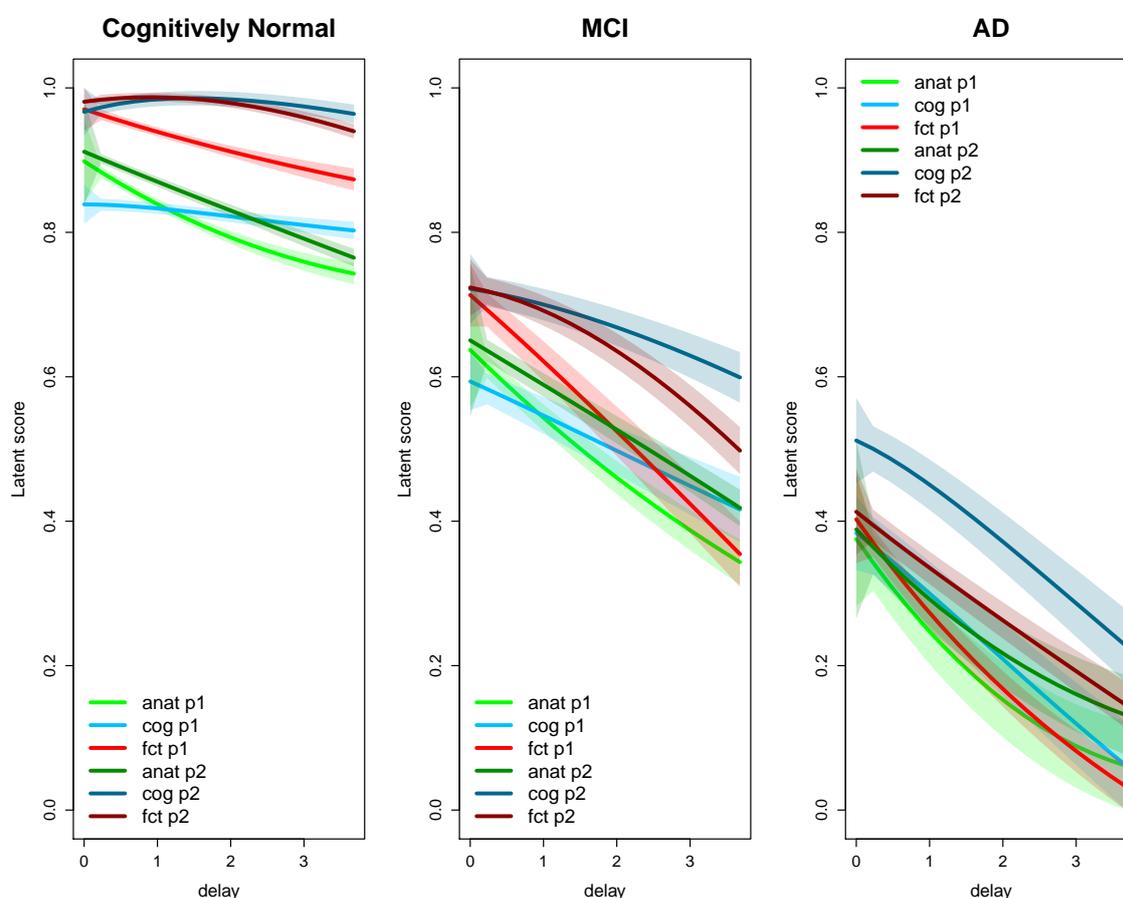


FIGURE 5.2 – Trajectoires moyennes prédites avec intervalles de confiance des dimensions anatomique, cognitive, fonctionnelle selon le stade clinique (CN, MCI, AD) pour deux profils de sujets : profil p1 = femme porteuse de l'allèle apolipoprotéine ϵ 4 ayant fait moins de 12 ans d'étude ; profil p2 = femme non porteuse de l'allèle apolipoprotéine ϵ 4 ayant fait plus de 12 ans d'étude.

5.2.1.3 Estimations des paramètres des modèles d'observation

Les estimations des paramètres des modèles d'observation des marqueurs sont données en annexes dans le Tableau A 6.1. Les fonctions de lien estimées sont représentées dans la Figure 5.1 sur la colonne de droite, les colonnes de gauche représentant les distributions observées des Z-scores avant et après transformation. Comme attendu, seule la fonction de lien du score composite fonctionnel montre un non-linéarité important en lien avec l'écart majeur à l'hypothèse de normalité.

5.2.1.4 Estimations des paramètres de la matrice d'influences temporelles entre les dimensions

Les estimations des paramètres d'influences temporelles entre les dimensions sont résumées dans le Tableau 5.3 et la Figure 5.3 fournit un graphe résumé de ces résultats. Dans ce graphe d'influences temporelles, nous avons représenté les arcs orientés d'une dimension vers une autre dimension au sein d'un stade lorsque le paramètre d'influence associé (combinaison de l'effet simple et de l'interaction avec le stade) avait une p-valeur au test de Wald multivarié inférieure à 10%.

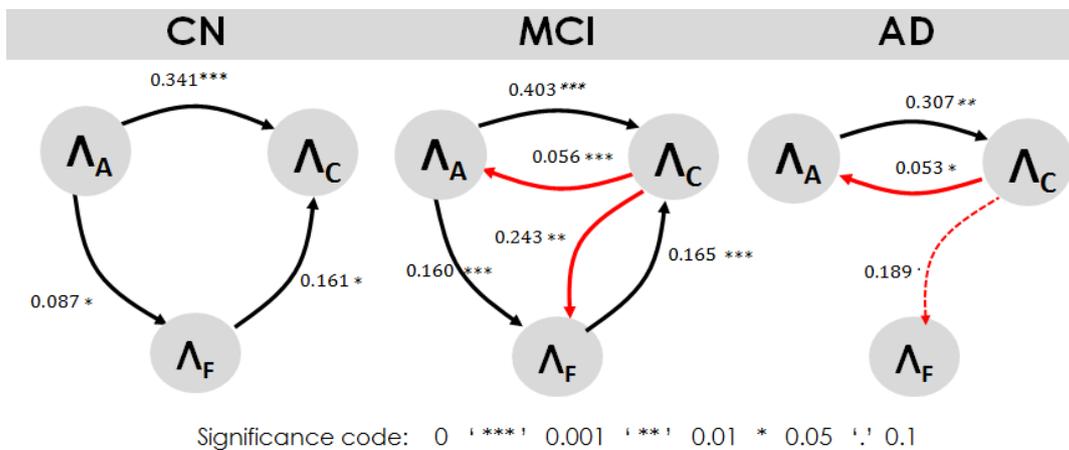


FIGURE 5.3 – Structure d'influences temporelles entre les dimensions anatomique (Λ_A), cognitive (Λ_C) et fonctionnelle (Λ_F) aux trois stades cliniques de la maladie d'Alzheimer (CN, MCI, AD). Une flèche de Λ_{d1} vers Λ_{d2} indique l'influence temporelle de la dimension $d1$ au temps t sur le taux de variation de la dimension Λ_{d2} au temps $t + \delta$ après ajustement sur le niveau des autres dimensions au temps t et sur les variables explicatives des trajectoires des dimensions; δ étant le pas de discrétisation.

Parmi les sujets contrôles, la dimension anatomique influence simultanément le taux de changement des dimensions cognitive et fonctionnelle. Un niveau anatomique plus grand

TABLE 5.3 – Estimations des paramètres de la matrice d’influences temporelles reliant les taux de variation sur un intervalle d’amplitude $\delta = 0.23 \approx 3$ mois des dimensions anatomique, cognitive et fonctionnelle à l’état précédent de ces dimensions.

| | Paramètre | Estimation | se | p-valeur | |
|--|-----------|-----------------|--------|----------|---------|
| <u>Influence sur la dim. anatomique :</u> | | | | | |
| Effet de la dim. anatomique | Intercept | α_{11}^0 | -0.086 | 0.023 | < 0.001 |
| | Intercept | α_{12}^0 | 0.008 | 0.014 | 0.577 |
| Effet de la dim. cognitive | MCI | α_{12}^1 | 0.048 | 0.017 | 0.004 |
| | AD | α_{12}^2 | 0.045 | 0.024 | 0.060 |
| Effet de la dim. fonctionnelle | Intercept | α_{13}^0 | 0.016 | 0.024 | 0.497 |
| | MCI | α_{13}^1 | -0.018 | 0.025 | 0.459 |
| | AD | α_{13}^2 | -0.016 | 0.030 | 0.588 |
| <u>Influence sur la dim. cognitive :</u> | | | | | |
| Effet de la dim. anatomique | Intercept | α_{21}^0 | 0.341 | 0.083 | < 0.001 |
| | MCI | α_{21}^1 | 0.062 | 0.054 | 0.252 |
| | AD | α_{21}^2 | -0.034 | 0.068 | 0.617 |
| Effet de la dim. cognitive | Intercept | α_{22}^0 | -0.670 | 0.151 | < 0.001 |
| | Intercept | α_{23}^0 | 0.161 | 0.082 | 0.048 |
| Effet de la dim. fonctionnelle | MCI | α_{23}^1 | 0.004 | 0.080 | 0.965 |
| | AD | α_{23}^2 | -0.059 | 0.092 | 0.519 |
| <u>Influence sur la dim. fonctionnelle :</u> | | | | | |
| Effet de la dim. anatomique | Intercept | α_{31}^0 | 0.087 | 0.043 | 0.043 |
| | MCI | α_{31}^1 | 0.072 | 0.054 | 0.179 |
| | AD | α_{31}^2 | 0.010 | 0.071 | 0.890 |
| Effet de la dim. cognitive | Intercept | α_{32}^0 | 0.081 | 0.058 | 0.161 |
| | MCI | α_{32}^1 | 0.162 | 0.060 | 0.007 |
| Effet de la dim. fonctionnelle | AD | α_{32}^2 | 0.108 | 0.073 | 0.140 |
| | Intercept | α_{33}^0 | -0.262 | 0.157 | 0.094 |

est associé à une augmentation des taux de variation des dimensions cognitive et fonctionnelle (c’est-à-dire à une déclin moindre). La dimension fonctionnelle est aussi associée au changement de la dimension cognitive avec un niveau fonctionnel plus important associé à un taux de variation augmenté de la cognition. Mais il n’y a ni de relation de la cognition vers le fonctionnel ni de relation directe du fonctionnel vers l’anatomique.

A partir du stade MCI, le niveau cognitif commence à devenir central dans les relations entre les dimensions. En effet, au stade MCI, en plus des relations précédemment trouvées au stade CN s’ajoutent des influences de la cognition vers les deux autres dimensions. Au stade AD, si les liens entre la dimension anatomique et la dimension cognitive restent, les liens avec la dimension fonctionnelle ne sont plus aussi présents.

5.2.1.5 Évaluation de l'adéquation du modèle aux données et analyses de sensibilité

Nous avons vérifié l'adéquation du modèle aux données en comparant les observations aux visites du suivi aux prédictions individuelles à ces mêmes visites. La figure 5.4 représente dans l'échelle transformée des données la moyenne des observations à chaque temps d'observation en trait plein et la moyenne des prédictions spécifiques aux individus (croix). Ce graphique illustre que le modèle prédit correctement les observations.

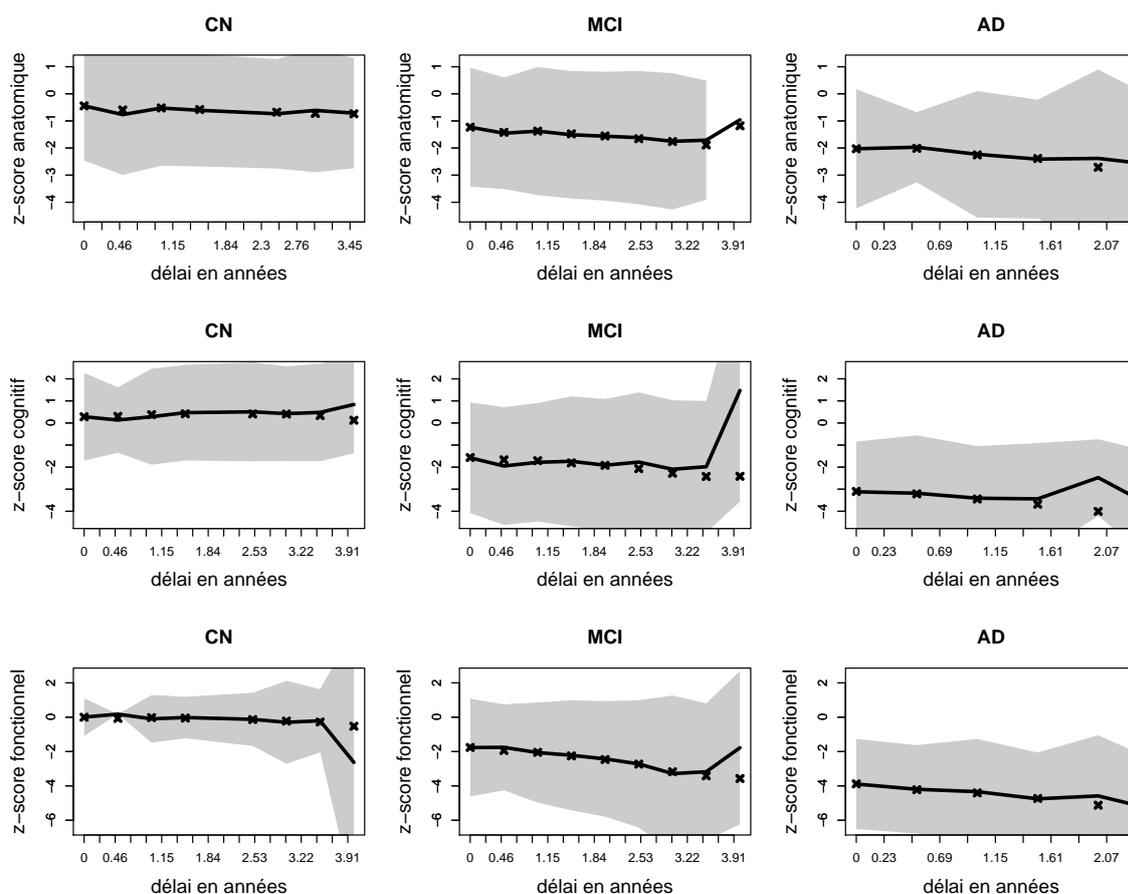


FIGURE 5.4 – Comparaison dans chaque stade clinique de la moyenne des observations transformées à chaque temps d'observation discret à la moyenne des prédictions.

5.2.2 Analyse de sensibilité

Dans cette section, nous avons effectué des analyses complémentaires pour évaluer la stabilité de la structure d'influences temporelles entre les dimensions anatomique, cognitive et fonctionnelle mesurées à partir des trois marqueurs composites généraux. D'après

les études de simulation effectuées dans le chapitre 3, il ressort qu'une mauvaise spécification du modèle peut impacter considérablement les paramètres du modèle et en particulier les paramètres de la structure d'influences temporelles entre les dimensions qui deviennent biaisés. Nous avons donc cherché à comparer les structures d'influences temporelles obtenues à partir de plusieurs spécifications de modèles.

Précisément, nous avons estimé quatre autres modèles et comparé leur structure d'influences temporelles à celle obtenue dans la section 5.2 (modèle 1) :

- modèle 2 : nous avons supposé que le taux de variation des dimensions n'était pas dépendant du temps (temps et interactions entre variables explicatives et temps éliminées du modèle pour le taux de variation) ;
- modèle 3 : nous avons supposé que les dimensions étaient indépendantes à l'inclusion (intercepts aléatoires non corrélées) ;
- modèles 4 et 5 : nous avons réduit le pas de discrétisation des dimensions à un pas de $\delta/2 = 0.115$ (modèle 4) et $\delta/3 = 0.083$ (modèle 5).

Nous n'avons pas directement comparé les estimations des α de la matrice d'influences mais les influences spécifiques au groupe (en combinant l'effet simple $\alpha_{..}^0$ et l'interaction avec le groupe $\alpha_{..}^k$, $k = 1, 2$). Les p-valeurs correspondantes ont été déduites du test de Wald multiple. Ce sont ces types d'influences spécifiques qui sont représentées dans le graphe résumé (Figure 5.3 pour le modèle 1).

Le tableau 5.4 résume les paramètres estimés d'influences temporelles par stade clinique pour les 5 modèles. Lorsque l'on ne modélise pas une évolution possible en fonction du temps du taux de variation, les estimations sont globalement peu changées. Les exceptions sont une petite atténuation de l'intensité de l'influence de l'anatomique vers le cognitif aux stades CN et AD. Il faut noter que dans le modèle 1, mis à part une accélération du déclin parmi les MCI, la composante temporelle du taux de variation ne semblait pas primordiale ; la considération du temps dans le taux de variation conduisait à un AIC de 73.62 points de plus que l'AIC du modèle avec un taux de variation constant des processus latent

Lorsque l'on ne tient pas compte de la corrélation entre les dimensions à l'inclusion, les conclusions restent aussi globalement les mêmes, mis à part l'influence entre le fonctionnel et le cognitif au stade AD qui devient significatif (comme cela était le cas aux stades précédents).

Lorsque l'on considère des pas de discrétisation plus fins, alors aucune différence n'est observée.

APPLICATIONS

TABLE 5.4 – Tableau comparatif des estimations (valeur) des éléments de la matrice d’influences temporelles et de leur p-valeur (p) pour cinq spécifications de modèles.

| Groupe | | modèle 1 | | modèle 2 | | modèle 3 | | modèle 4 | | modèle 5 | |
|--------|----------|----------|--------|----------|--------|----------|--------|----------|--------|----------|--------|
| | | valeur | p |
| CN | a_{11} | -0.086 | <0.001 | -0.090 | <0.001 | -0.111 | <0.001 | -0.088 | <0.001 | -0.088 | <0.001 |
| | a_{12} | 0.008 | 0.577 | 0.018 | 0.215 | 0.011 | 0.447 | 0.007 | 0.635 | 0.007 | 0.616 |
| | a_{13} | 0.016 | 0.497 | 0.013 | 0.587 | 0.012 | 0.615 | 0.014 | 0.550 | 0.013 | 0.590 |
| | a_{21} | 0.341 | <0.001 | 0.140 | 0.001 | 0.272 | <0.001 | 0.386 | <0.001 | 0.395 | <0.001 |
| | a_{22} | -0.670 | <0.001 | -0.610 | <0.001 | -0.493 | <0.001 | -0.765 | <0.001 | -0.779 | <0.001 |
| | a_{23} | 0.161 | 0.048 | 0.178 | 0.013 | 0.156 | 0.016 | 0.177 | 0.421 | 0.171 | 0.062 |
| | a_{31} | 0.087 | 0.043 | 0.083 | 0.045 | 0.077 | 0.032 | 0.093 | 0.043 | 0.097 | 0.041 |
| | a_{32} | 0.081 | 0.161 | 0.069 | 0.137 | 0.031 | 0.442 | 0.090 | 0.237 | 0.096 | 0.132 |
| | a_{33} | -0.262 | 0.094 | -0.315 | 0.015 | -0.095 | 0.130 | -0.301 | 0.077 | -0.322 | 0.054 |
| MCI | a_{11} | -0.086 | <0.001 | -0.090 | <0.001 | -0.111 | <0.001 | -0.088 | <0.001 | -0.088 | <0.001 |
| | a_{12} | 0.056 | <0.001 | 0.066 | <0.001 | 0.059 | <0.001 | 0.055 | <0.001 | 0.056 | <0.001 |
| | a_{13} | -0.003 | 0.959 | -0.003 | 0.952 | -0.006 | 0.790 | -0.003 | 0.957 | -0.003 | 0.952 |
| | a_{21} | 0.403 | <0.001 | 0.201 | <0.001 | 0.313 | <0.001 | 0.453 | <0.001 | 0.463 | <0.001 |
| | a_{22} | -0.670 | <0.001 | -0.610 | <0.001 | -0.493 | <0.001 | -0.765 | <0.001 | -0.779 | <0.001 |
| | a_{23} | 0.165 | <0.001 | 0.179 | <0.001 | 0.163 | <0.001 | 0.178 | <0.001 | 0.182 | <0.001 |
| | a_{31} | 0.160 | 0.001 | 0.158 | <0.001 | 0.143 | <0.001 | 0.169 | 0.002 | 0.177 | 0.002 |
| | a_{32} | 0.243 | 0.005 | 0.240 | <0.001 | 0.163 | <0.001 | 0.264 | 0.005 | 0.273 | 0.003 |
| | a_{33} | -0.262 | 0.094 | -0.315 | 0.015 | -0.095 | 0.130 | -0.301 | 0.077 | -0.322 | 0.054 |
| AD | a_{11} | -0.086 | <0.001 | -0.090 | <0.001 | -0.111 | <0.001 | -0.088 | <0.001 | -0.088 | <0.001 |
| | a_{12} | 0.053 | 0.029 | 0.068 | 0.005 | 0.053 | 0.039 | 0.053 | 0.030 | 0.053 | 0.035 |
| | a_{13} | -0.000 | 1.000 | 0.001 | 1.000 | -0.009 | 0.903 | -0.003 | 0.989 | -0.001 | 0.998 |
| | a_{21} | 0.307 | 0.002 | 0.111 | 0.092 | 0.231 | 0.001 | 0.351 | 0.002 | 0.355 | 0.002 |
| | a_{22} | -0.670 | <0.001 | -0.610 | <0.001 | -0.493 | <0.001 | -0.765 | <0.001 | -0.779 | <0.001 |
| | a_{23} | 0.102 | 0.169 | 0.136 | 0.035 | 0.146 | 0.012 | 0.110 | 0.190 | 0.112 | 0.188 |
| | a_{31} | 0.097 | 0.293 | 0.098 | 0.243 | 0.077 | 0.322 | 0.098 | 0.331 | 0.102 | 0.317 |
| | a_{32} | 0.189 | 0.058 | 0.182 | 0.016 | 0.112 | 0.121 | 0.202 | 0.059 | 0.211 | 0.052 |
| | a_{33} | -0.262 | 0.094 | -0.315 | 0.015 | -0.095 | 0.130 | -0.301 | 0.077 | -0.322 | 0.054 |

5.2.3 Conclusion

Dans cette application, nous avons exploré les relations temporelles entre les dimensions anatomique, cognitive et fonctionnelle en supposant des marqueurs très globaux de ces dimensions.

Nous avons trouvé des influences temporelles de la dimension anatomique vers les dimensions cognitive et fonctionnelle et de la dimension fonctionnelle vers la dimension cognitive dans le stade CN.

A partir du stade MCI, on trouve un retour de la dimension cognitive vers la dimension anatomique ce qui suggère des inter-relations entre ces deux dimensions. Au stade MCI, la dimension cognitive impacte aussi la sphère fonctionnelle alors que cette association n'est pas retrouvée au stade AD.

Les résultats dans le stade AD sont à prendre avec précautions à cause de peu de données dans ce stade. L'influence temporelle de la dimension fonctionnelle sur la dimension cognitive trouvée peut paraître contre-intuitive à première vue. Mais elle peut être expliquée par le fait que la FAQ, qui mesure la dimension fonctionnelle, n'inclut pas que des items sur la dépendance mais aussi sur les activités sociales qui peuvent être très en amont de la maladie et expliquer l'influence de la dimension fonctionnelle sur la dimension cognitive.

Enfin, les analyses de sensibilité montrent dans l'ensemble une bonne consistance des résultats.

5.3 Application 2 : Exploration plus fine des dimensions anatomique et cognitive

5.3.0.1 Objectif

L'objectif de cette deuxième application est de reprendre l'approche de modélisation de l'application 1 tout en définissant plus finement les dimensions anatomique et cognitive pour affiner les relations temporelles entre les dimensions anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer. Dans un premier temps, nous avons recentré la dimension anatomique en nous focalisant sur le volume relatif de l'hippocampe qui définit la dimension « atrophie de l'hippocampe », ensuite nous avons recentré la dimension cognitive en considérant uniquement deux fonctions cognitives a priori pertinentes pour la maladie d'Alzheimer : la « mémoire épisodique » et la « fluence verbale ». Enfin nous avons modélisé conjointement les 4 dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction » et leur structure d'influences temporelles dans les trois stades CN, MCI et AD.

5.3.0.2 Spécification des modèles

Par rapport à l'application 1, nous avons considéré dans cette deuxième application les spécifications suivantes :

- Le taux de variation des dimensions au cours du temps est constant sur l'intervalle de temps t et $t + \delta$. Ce choix est justifié par les résultats de l'analyse de sensibilité

faite à la section 5.2.2 et par le pas de discrétisation petit $\delta = 0.23$.

- Un intercept aléatoire sur le niveau initial de chaque dimension.
- Un intercept aléatoire sur le taux de variation de chaque dimension.

5.3.1 Affinement de la dimension anatomique

Ici nous avons décidé d'affiner la dimension anatomique en nous focalisant sur l'atrophie de l'hippocampe. L'hippocampe est une structure du cerveau située dans le lobe temporal médian, sous la surface du cortex. Il joue un rôle central dans la mémoire [Mormino et al., 2008]. Nous avons utilisé la moyenne standardisée des volumes relatifs (au volume total du cerveau) des hippocampes gauche et droite comme mesure de l'atrophie de l'hippocampe. La figure 5.5 présente le schéma de modélisation qui est traduit en équation 5.2.

$$\left\{ \begin{array}{l}
 \Lambda_{i1}(0) = \beta_{01} + \beta_{11}\mathbf{MCI}_i + \beta_{21}\mathbf{AD}_i + \beta_{31}\mathbf{AGE0}_i \\
 \quad \quad \quad + \beta_{41}\mathbf{gender}_i + \beta_{51}\mathbf{educ12}_i + \beta_{61}\mathbf{apoe4}_i + u_{i1} \\
 \Lambda_{i2}(0) = \beta_{02} + \beta_{12}\mathbf{MCI}_i + \beta_{22}\mathbf{AD}_i + \beta_{32}\mathbf{AGE0}_i \\
 \quad \quad \quad + \beta_{42}\mathbf{gender}_i + \beta_{52}\mathbf{educ12}_i + \beta_{62}\mathbf{apoe4}_i + u_{i2} \\
 \Lambda_{i3}(0) = \beta_{03} + \beta_{13}\mathbf{MCI}_i + \beta_{23}\mathbf{AD}_i + \beta_{33}\mathbf{AGE0}_i + \beta_{43}\mathbf{apoe4}_i \\
 \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}\mathbf{MCI}_i + \gamma_{21}\mathbf{AD}_i + \gamma_{31}\mathbf{gender}_i + \gamma_{41}\mathbf{educ12}_i + \gamma_{51}\mathbf{apoe4}_i + v_{i1} \\
 \quad \quad \quad + \alpha_{11}^0\Lambda_{i1}(t_{ij} - \delta) + (\alpha_{12}^0 + \alpha_{12}^1\mathbf{MCI} + \alpha_{12}^2\mathbf{AD})\Lambda_{i2}(t_{ij} - \delta) \\
 \quad \quad \quad + (\alpha_{13}^0 + \alpha_{13}^1\mathbf{MCI} + \alpha_{13}^2\mathbf{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}\mathbf{MCI}_i + \gamma_{22}\mathbf{AD}_i + \gamma_{32}\mathbf{gender}_i + \gamma_{42}\mathbf{educ12}_i + \gamma_{52}\mathbf{apoe4}_i + v_{i2} \\
 \quad \quad \quad + \alpha_{22}^0\Lambda_{i2}(t_{ij} - \delta) + (\alpha_{21}^0 + \alpha_{21}^1\mathbf{MCI} + \alpha_{21}^2\mathbf{AD})\Lambda_{i1}(t_{ij} - \delta) \\
 \quad \quad \quad + (\alpha_{31}^0 + \alpha_{31}^1\mathbf{MCI} + \alpha_{31}^2\mathbf{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i3}(t_{ij})}{\delta} = \gamma_{03} + \gamma_{13}\mathbf{MCI}_i + \gamma_{23}\mathbf{AD}_i + \gamma_{33}\mathbf{apoe4}_i + v_{i3} \\
 \quad \quad \quad + \alpha_{33}^0\Lambda_{i3}(t_{ij} - \delta) + (\alpha_{31}^0 + \alpha_{31}^1\mathbf{MCI} + \alpha_{31}^2\mathbf{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \quad \quad \quad + (\alpha_{32}^0 + \alpha_{32}^1\mathbf{MCI} + \alpha_{32}^2\mathbf{AD})\Lambda_{i2}(t_{ij} - \delta)
 \end{array} \right. \tag{5.2}$$

avec $\Lambda_{i1}(t_{ij})$, $\Lambda_{i2}(t_{ij})$, $\Lambda_{i3}(t_{ij})$ respectivement les dimensions « atrophie de l'hippocampe », « cognition globale », « fonctionnelle » chez le sujet i au temps t_{ij} , $\Delta\Lambda_i = \Lambda_i(t_{ij}) - \Lambda_i(t_{ij} - \delta)$, et δ le pas de discrétisation ; Les vecteurs β , α et γ incluent les paramètres d'effets fixes (les paramètres β_0 sont fixés à 0 par contrainte d'identifiabilité), $(u_{i1}, u_{i2}, u_{i3}, v_{i1}, v_{i2}, v_{i3})^\top$ un vecteur d'effets aléatoires centrés de matrice de variance-covariance \mathbf{B} . \mathbf{B} suit la structure indiquée en méthodologie : les intercepts aléatoires sur le niveau initial sont standardisées, les effets aléatoires sont corrélés au sein de chaque dimension latente. Par contre, les effets aléatoires entre dimensions sont indépendants sauf les intercepts aléatoires sur le niveau initial qui captent une potentielle corrélation à l'inclusion.

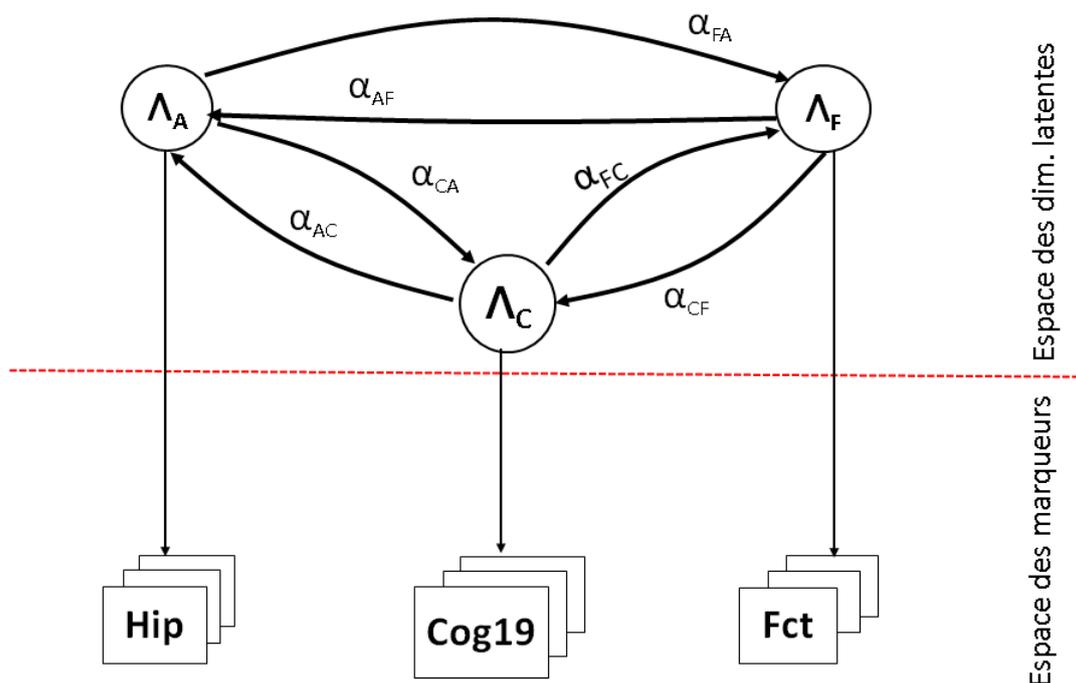


FIGURE 5.5 – Schéma de modélisation des dimensions « atrophie de l’hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l’hippocampe (Hip), « cognition globale » (Λ_C) mesurée par un composite score de 19 scores cognitifs standardisés (Cog19), dimension fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (Fct).

5.3.1.1 Résultats de la modélisation

Le tableau 5.5 résume les estimations des paramètres fixes du modèle structurel et la figure 5.6 résume les influences temporelles absolues entre les trois dimensions d’intérêts.

TABLE 5.5 – Estimations des paramètres fixes du modèle conjoint des dimensions « atrophie de l’hippocampe », « cognition globale » et « fonction ».

| L = -12943.81, AIC = 26051.61, p = 82 | | | |
|---------------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| Niveau initial | | | |
| β_{11} | -1.012 | 0.101 | < 0.001 |
| β_{21} | -1.689 | 0.128 | < 0.001 |
| β_{31} | -0.534 | 0.062 | < 0.001 |
| β_{41} | -0.359 | 0.080 | < 0.001 |

APPLICATIONS

Suite de la Table 5.5 - Estimations des paramètres du modèle conjoint des dimensions « atrophie de l'hippocampe », « cognition globale » et « fonction ».

| L = -12943.81, AIC = 26051.61, p = 82 | | | |
|---------------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| β_{51} | -0.113 | 0.104 | 0.277 |
| β_{61} | -0.364 | 0.085 | < 0.001 |
| β_{12} | -1.635 | 0.116 | < 0.001 |
| β_{22} | -2.999 | 0.158 | < 0.001 |
| β_{32} | -0.275 | 0.058 | < 0.001 |
| β_{42} | -0.180 | 0.084 | 0.033 |
| β_{52} | 0.518 | 0.112 | < 0.001 |
| β_{62} | -0.352 | 0.091 | < 0.001 |
| β_{13} | -1.757 | 0.132 | < 0.001 |
| β_{23} | -3.962 | 0.201 | < 0.001 |
| β_{33} | -0.035 | 0.070 | 0.615 |
| β_{43} | -0.136 | 0.101 | 0.175 |
| Taux de variation | | | |
| γ_{01} | -0.086 | 0.015 | < 0.001 |
| γ_{11} | -0.022 | 0.019 | 0.243 |
| γ_{21} | -0.120 | 0.050 | 0.016 |
| γ_{31} | 0.031 | 0.010 | 0.002 |
| γ_{41} | -0.010 | 0.012 | 0.422 |
| γ_{51} | -0.051 | 0.010 | < 0.001 |
| γ_{02} | 0.065 | 0.058 | 0.264 |
| γ_{12} | -0.390 | 0.116 | 0.001 |
| γ_{22} | -1.015 | 0.256 | < 0.001 |
| γ_{32} | 0.042 | 0.041 | 0.311 |
| γ_{42} | 0.204 | 0.072 | 0.004 |
| γ_{52} | -0.166 | 0.048 | 0.001 |
| γ_{03} | -0.054 | 0.046 | 0.245 |
| γ_{13} | 0.003 | 0.104 | 0.979 |
| γ_{23} | -0.516 | 0.263 | 0.050 |
| γ_{33} | -0.119 | 0.043 | 0.006 |
| Influences temporelles | | | |
| α_{11}^0 | -0.006 | 0.012 | 0.604 |
| α_{12}^0 | 0.002 | 0.009 | 0.804 |

Suite de la Table 5.5 - Estimations des paramètres du modèle conjoint des dimensions « atrophie de l'hippocampe », « cognition globale » et « fonction ».

| L = -12943.81, AIC = 26051.61, p = 82 | | | |
|---------------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| α_{12}^1 | 0.025 | 0.010 | 0.018 |
| α_{12}^2 | 0.002 | 0.014 | 0.915 |
| α_{13}^0 | 0.018 | 0.014 | 0.188 |
| α_{13}^1 | -0.006 | 0.014 | 0.690 |
| α_{13}^2 | -0.012 | 0.017 | 0.474 |
| α_{21}^0 | 0.164 | 0.042 | < 0.001 |
| α_{21}^1 | -0.030 | 0.038 | 0.420 |
| α_{21}^2 | -0.141 | 0.053 | 0.008 |
| α_{22}^0 | -0.313 | 0.079 | < 0.001 |
| α_{23}^0 | 0.098 | 0.054 | 0.073 |
| α_{23}^1 | 0.037 | 0.055 | 0.504 |
| α_{23}^2 | 0.022 | 0.065 | 0.730 |
| α_{31}^0 | 0.068 | 0.036 | 0.059 |
| α_{31}^1 | 0.082 | 0.043 | 0.055 |
| α_{31}^2 | -0.140 | 0.054 | 0.010 |
| α_{32}^0 | 0.045 | 0.040 | 0.260 |
| α_{32}^1 | 0.128 | 0.046 | 0.005 |
| α_{32}^2 | 0.121 | 0.060 | 0.043 |
| α_{33}^0 | -0.075 | 0.067 | 0.261 |
| α_{33}^0 | -0.075 | 0.064 | 0.245 |

Interprétation des paramètres de trajectoires des dimensions

D'après les estimations des paramètres du niveau initial et du taux de variation des dimensions données dans le tableau (voir tableau 5.5), on voit que, par rapport à un sujet de référence (c'est-à-dire cognitivement normal (CN), âgé de 75.4 ans à l'inclusion, de sexe féminin, de bas niveau d'étude et non porteur de l'allèle apolipoprotéine $\epsilon 4$), un sujet au stade MCI ou AD, ou plus âgé à l'inclusion ou de sexe masculin ou porteur de l'allèle apolipoprotéine $\epsilon 4$ a un volume relatif de l'hippocampe plus petit à l'inclusion.

Un sujet au stade MCI ou AD ou plus âgé à l'inclusion ou de sexe masculin ou porteur de l'allèle apolipoprotéine $\epsilon 4$ a un niveau cognitif global plus bas que le sujet de référence à l'inclusion. Tandis qu'un sujet de haut niveau d'étude à un niveau cognitif plus élevé que le sujet de référence à l'inclusion.

APPLICATIONS

Seuls les sujets au stade MCI ou AD ont un niveau fonctionnel plus bas que le sujet de référence à l'inclusion.

S'agissant de la dynamique des dimensions au cours du temps, ajusté sur le niveau courant du système, le volume relatif de l'hippocampe du sujet de référence diminue au cours du temps. Relativement au déclin du volume relatif de l'hippocampe chez le sujet de référence, un sujet au stade AD ou porteur de l'allèle apolipoprotéine $\epsilon 4$ a un déclin plus prononcé du volume relatif de l'hippocampe. Un sujet de sexe masculin a un déclin moins prononcé du volume relatif de l'hippocampe relativement au sujet de référence.

La dimension cognitive globale semble ne pas varier au cours du temps chez le sujet de référence après ajustement sur le niveau du système. Par contre, elle décline chez un sujet au stade MCI ou AD ou porteur de l'allèle apolipoprotéine $\epsilon 4$ relativement au sujet de référence. Ce déclin est moins prononcé chez un sujet de haut niveau d'étude relativement au sujet de référence.

Il y a un déclin significatif de la dimension fonctionnelle chez un sujet au stade AD ou porteur de l'allèle apolipoprotéine $\epsilon 4$ relativement au sujet de référence ajusté sur le niveau courant du système.

Ces résultats sont bien ceux attendus.

Interprétation des paramètres d'influences temporelles

Sur la base des influences temporelles entre les dimensions représentées dans la figure 5.6, chez un sujet cognitivement normal on trouve principalement qu'un niveau plus bas du volume relatif de l'hippocampe est associé à un déclin de la dimension cognitive globale après ajustement sur le niveau fonctionnel courant.

Chez un sujet au stade MCI, et relativement à un sujet cognitivement normal (CN), un niveau plus faible de la dimension cognitive globale est associé à une augmentation du déclin du volume relatif de l'hippocampe après ajustement sur le niveau courant de la dimension fonctionnelle. Un niveau plus faible de la dimension cognitive globale est associé à une augmentation du déclin de la dimension fonctionnelle après ajustement sur le niveau courant du volume relatif de l'hippocampe.

Chez un sujet au stade AD, et relativement à un sujet cognitivement normal (CN), un niveau plus faible de la dimension cognitive globale est aussi (comme en MCI) associé à une augmentation du déclin de la dimension fonctionnelle après ajustement sur le niveau courant du volume relatif de l'hippocampe. Par contre, l'association entre le volume relatif de l'hippocampe et le déclin cognitif disparaît.

La figure 5.6 résume les influences absolues trouvées entre le volume relatif de l'hippocampe, la dimensions cognitive globale et la dimension fonctionnelle dans les trois stade cliniques de la

maladie d'Alzheimer. Globalement, ces résultats sont assez concordants avec ceux trouvés dans l'application 1.

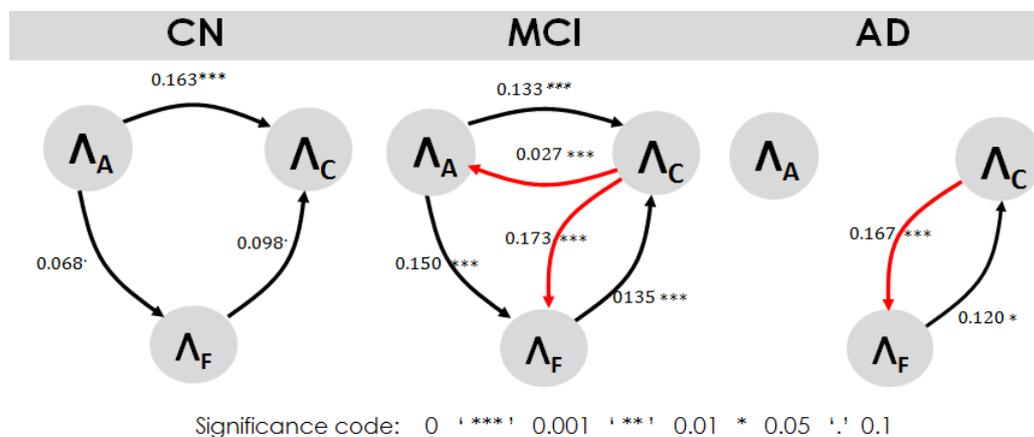


FIGURE 5.6 – Influences temporelles entre les dimensions « atrophie de l'hippocampe » (Λ_A), « cognitive globale » (Λ_C) et fonctionnelle (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer).

5.3.2 Affinement des dimensions anatomique et cognitive

Ici nous avons décidé de continuer la démarche de modélisation plus fine des dimensions en investiguant les relations entre la dimension « atrophie de l'hippocampe » mesurée par le volume relatif standardisé de l'hippocampe (*Hip*), la dimension fonctionnelle mesurée par la somme standardisée des scores aux 30 items du test « Functionary Assessment Questionnaire » (*Fct*) et la dimension « cognition » définie par deux sous-fonctions : un score de mémoire épisodique et un score de fluence verbale. Ces deux sous-dimensions sont connues dans la littérature comme affectées par le processus de vieillissement et la démence de type Alzheimer (en particulier pour la mémoire épisodique) [Gomez et White, 2006; Ricci et al., 2012].

Pour la mémoire épisodique nous avons utilisé les scores répétés au test de « Auditory Verbal Learning Test » [Rey, 1964]. Il s'agit d'un test psychométrique d'apprentissage verbal qui se base sur une liste de mots pour détecter les atteintes de la mémoire épisodique dans la maladie d'Alzheimer et les démences apparentées. Nous avons construit un composite score répété (*Avlt*) à partir des sous-scores standardisés « du rappel immédiat », « du rappel différé » et de « la reconnaissance différée de mots ».

Pour la fluence verbale, nous avons utilisé les scores répétés au test « Category Fluency Test » qui consiste à demander au participant de citer, en un temps donné, le plus grand nombre de mots appartenant à une catégorie sémantique [Caramelli et al., 2007]. La fluence verbale mesure de façon plus globale le langage. Nous avons utilisé le score standardisé répété au test dans la catégorie lexicale « animale » (*CatFlu*).

APPLICATIONS

La figure 5.7 présente le schéma de modélisation. Le modèle structurel est le même que celui défini en équation 5.2 dans laquelle $\Lambda_{i2}(t_{ij})$ représente la dimension « cognition globale » définie de façon plus précise.

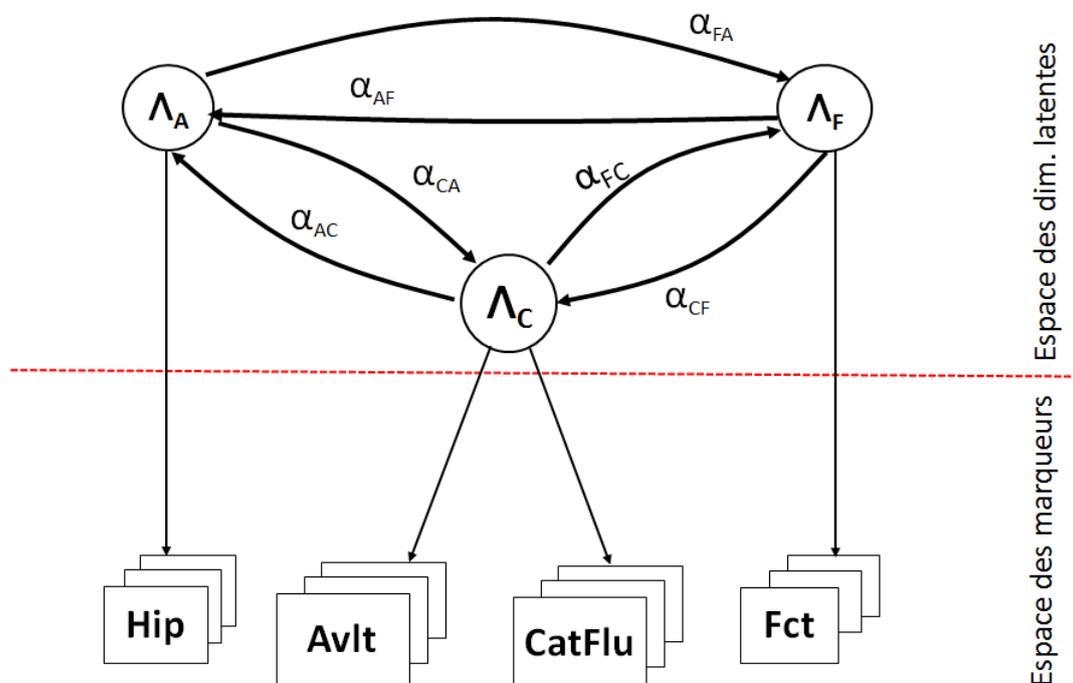


FIGURE 5.7 – Schéma de modélisation des dimensions « atrophie de l'hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l'hippocampe (Hip), « cognition » (Λ_C) mesurée à la fois par le composite score de sous-scores standardisés du test « Auditory Verbal Learning Test » (Avlt) et par le score standardisé au test « Category Fluency Test » dans le champ lexical animal (CatFlu), fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (Fct).

5.3.2.1 Résultats de la modélisation

Le tableau 5.6 résume les estimations des paramètres fixes du modèle structurel. Comparativement aux résultats du modèle de la section 5.3.1, les conclusions relatives aux niveaux initiaux des dimensions, leur taux de variation et leur structure d'influences temporelles restent les mêmes lorsque l'on se focalise sur deux scores de cognition particulièrement pertinents.

La figure 5.8 résume les influences absolues entre le volume relatif de l'hippocampe, la dimension cognition globale et la dimension fonctionnelle dans les trois stades cliniques de la maladie d'Alzheimer.

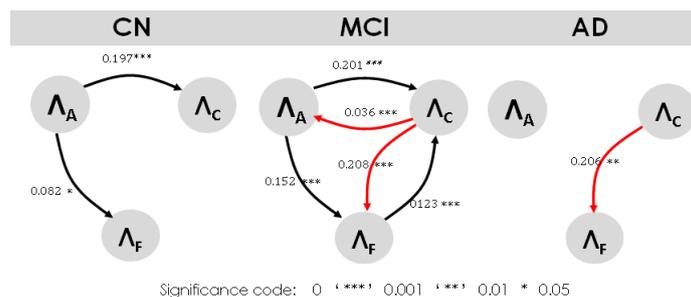


FIGURE 5.8 – Influences temporelles entre les dimensions « atrophie de l’hippocampe » (Λ_A), « cognition » (Λ_C) et fonction (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer).

TABLE 5.6 – Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l’hippocampe » et « cognition ».

| L = -19353.5, AIC = 38885, p = 89 | | | |
|-----------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| Niveau initial | | | |
| β_{11} | -1.014 | 0.101 | < 0.001 |
| β_{21} | -1.690 | 0.128 | < 0.001 |
| β_{31} | -0.529 | 0.062 | < 0.001 |
| β_{41} | -0.353 | 0.080 | < 0.001 |
| β_{51} | -0.114 | 0.104 | 0.276 |
| β_{61} | -0.364 | 0.085 | < 0.001 |
| β_{12} | -1.678 | 0.126 | < 0.001 |
| β_{22} | -2.859 | 0.170 | < 0.001 |
| β_{32} | -0.235 | 0.067 | < 0.001 |
| β_{42} | -0.216 | 0.092 | 0.019 |
| β_{52} | 0.346 | 0.121 | 0.004 |
| β_{62} | -0.326 | 0.098 | 0.001 |
| β_{13} | -1.637 | 0.128 | < 0.001 |
| β_{23} | -3.862 | 0.196 | < 0.001 |
| β_{33} | -0.040 | 0.071 | 0.577 |
| β_{43} | -0.139 | 0.099 | 0.161 |
| Taux de variation | | | |
| γ_{01} | -0.091 | 0.014 | < 0.001 |
| γ_{11} | -0.003 | 0.020 | 0.885 |
| γ_{21} | -0.104 | 0.050 | 0.038 |

APPLICATIONS

Suite de la Table 5.6 - Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe » et « cognition ».

| L = -19353.5, AIC = 38885, p = 89 | | | |
|-----------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| γ_{31} | 0.034 | 0.010 | 0.001 |
| γ_{41} | -0.007 | 0.012 | 0.566 |
| γ_{51} | -0.047 | 0.010 | < 0.001 |
| γ_{02} | 0.025 | 0.080 | 0.756 |
| γ_{12} | -0.504 | 0.153 | 0.001 |
| γ_{22} | -1.686 | 0.357 | < 0.001 |
| γ_{32} | 0.070 | 0.051 | 0.168 |
| γ_{42} | 0.162 | 0.079 | 0.039 |
| γ_{52} | -0.176 | 0.059 | 0.003 |
| γ_{03} | -0.044 | 0.047 | 0.347 |
| γ_{13} | -0.034 | 0.092 | 0.714 |
| γ_{23} | -0.685 | 0.258 | 0.008 |
| γ_{33} | -0.120 | 0.046 | 0.009 |
| Influences temporelles | | | |
| α_{11}^0 | -0.005 | 0.012 | 0.696 |
| α_{12}^0 | 0.012 | 0.008 | 0.124 |
| α_{12}^1 | 0.024 | 0.010 | 0.015 |
| α_{12}^2 | -0.002 | 0.014 | 0.861 |
| α_{13}^0 | 0.015 | 0.014 | 0.267 |
| α_{13}^1 | -0.007 | 0.015 | 0.644 |
| α_{13}^2 | -0.011 | 0.017 | 0.508 |
| α_{21}^0 | 0.197 | 0.055 | < 0.001 |
| α_{21}^1 | 0.004 | 0.045 | 0.925 |
| α_{21}^2 | -0.242 | 0.066 | < 0.001 |
| α_{22}^0 | -0.370 | 0.102 | < 0.001 |
| α_{23}^0 | 0.157 | 0.083 | 0.059 |
| α_{23}^1 | -0.034 | 0.082 | 0.677 |
| α_{23}^2 | -0.105 | 0.093 | 0.258 |
| α_{31}^0 | 0.082 | 0.040 | 0.038 |
| α_{31}^1 | 0.070 | 0.047 | 0.136 |
| α_{31}^2 | -0.165 | 0.059 | 0.005 |
| α_{32}^0 | 0.038 | 0.043 | 0.376 |
| α_{32}^1 | 0.170 | 0.053 | 0.001 |

Suite de la Table 5.6 - Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe » et « cognition ».

| L = -19353.5, AIC = 38885, p = 89 | | | |
|-----------------------------------|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| α_{32}^2 | 0.169 | 0.068 | 0.013 |
| α_{33}^0 | -0.144 | 0.065 | 0.028 |

5.3.3 Modélisation des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction »

Pour explorer encore plus finement les relations entre dimensions, nous avons séparé la dimension cognitive en deux sous-dimensions « mémoire épisodique » (Λ_M) mesurée par le composite score du test « Auditory Verbal Learning Test » (**Avl_t**) et « fluence verbale » (Λ_L) mesurée par le score standardisé au test « Category Fluency Test » pour la catégorie animale (**CatFlu**) en plus des dimensions « atrophie de l'hippocampe » et « fonction ».

La figure 5.9 représente le schéma de modélisation. Le modèle structurel estimé est de la même forme que le modèle de l'équation 5.2 sauf que la dimension cognitive a été éclatée en deux sous-dimensions. Le modèle structurel estimé est décrit en équation 5.3.

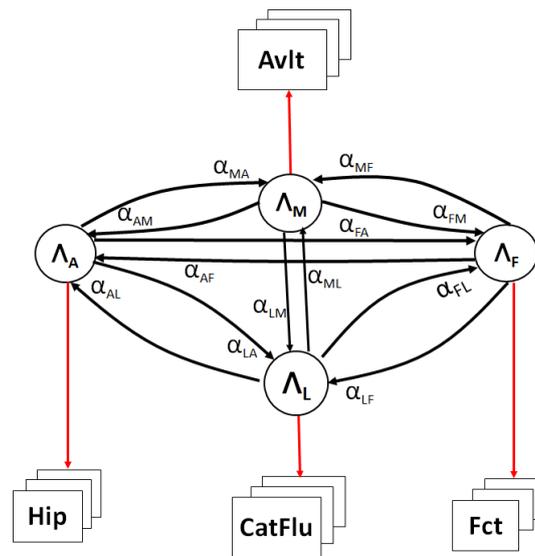


FIGURE 5.9 – Schéma de modélisation des dimensions « atrophie de l’hippocampe » (Λ_A) mesurée par le volume relatif standardisé de l’hippocampe (**Hip**), la dimension « mémoire épisodique » (Λ_M) mesurée par un composite score de sous-scores standardisés du tests « Auditory Verbal Learning Test » (**Avlt**), la dimension « fluence verbale » (Λ_L) mesurée par le score standardisé au test « Category Fluency Test » pour la catégorie animale (**CatFlu**) et la dimension fonctionnelle (Λ_F) mesurée par la somme standardisée des scores aux 30 items du « Functionary Assessment Questionnaire » (**Fct**).

$$\left\{ \begin{array}{l}
 \Lambda_{i1}(0) = \beta_{01} + \beta_{11}\text{MCI}_i + \beta_{21}\text{AD}_i + \beta_{31}\text{AGEO}_i \\
 \quad + \beta_{41}\text{gender}_i + \beta_{51}\text{educ12}_i + \beta_{61}\text{apoe4}_i + u_{i1} \\
 \Lambda_{i2}(0) = \beta_{02} + \beta_{12}\text{MCI}_i + \beta_{22}\text{AD}_i + \beta_{32}\text{AGEO}_i \\
 \quad + \beta_{42}\text{gender}_i + \beta_{52}\text{educ12}_i + \beta_{62}\text{apoe4}_i + u_{i2} \\
 \Lambda_{i3}(0) = \beta_{03} + \beta_{13}\text{MCI}_i + \beta_{23}\text{AD}_i + \beta_{33}\text{AGEO}_i \\
 \quad + \beta_{43}\text{gender}_i + \beta_{53}\text{educ12}_i + \beta_{63}\text{apoe4}_i + u_{i3} \\
 \Lambda_{i4}(0) = \beta_{04} + \beta_{14}\text{MCI}_i + \beta_{24}\text{AD}_i + \beta_{34}\text{AGEO}_i + \beta_{44}\text{apoe4}_i \\
 \frac{\Delta\Lambda_{i1}(t_{ij})}{\delta} = \gamma_{01} + \gamma_{11}\text{MCI}_i + \gamma_{21}\text{AD}_i + \gamma_{31}\text{gender}_i + \gamma_{41}\text{educ12}_i + \gamma_{51}\text{apoe4}_i + v_{i1} \\
 \quad + \alpha_{11}^0\Lambda_{i1}(t_{ij} - \delta) + (\alpha_{12}^0 + \alpha_{12}^1\text{MCI} + \alpha_{12}^2\text{AD})\Lambda_{i2}(t_{ij} - \delta) \\
 \quad + (\alpha_{13}^0 + \alpha_{13}^1\text{MCI} + \alpha_{13}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \quad + (\alpha_{14}^0 + \alpha_{14}^1\text{MCI} + \alpha_{14}^2\text{AD})\Lambda_{i4}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i2}(t_{ij})}{\delta} = \gamma_{02} + \gamma_{12}\text{MCI}_i + \gamma_{22}\text{AD}_i + \gamma_{32}\text{gender}_i + \gamma_{42}\text{educ12}_i + \gamma_{52}\text{apoe4}_i + v_{i2} \\
 \quad + \alpha_{22}^0\Lambda_{i2}(t_{ij} - \delta) + (\alpha_{21}^0 + \alpha_{21}^1\text{MCI} + \alpha_{21}^2\text{AD})\Lambda_{i1}(t_{ij} - \delta) \\
 \quad + (\alpha_{23}^0 + \alpha_{23}^1\text{MCI} + \alpha_{23}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta) \\
 \quad + (\alpha_{24}^0 + \alpha_{24}^1\text{MCI} + \alpha_{24}^2\text{AD})\Lambda_{i4}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i3}(t_{ij})}{\delta} = \gamma_{03} + \gamma_{13}\text{MCI}_i + \gamma_{23}\text{AD}_i + \gamma_{33}\text{gender}_i + \gamma_{43}\text{educ12}_i + \gamma_{53}\text{apoe4}_i + v_{i3} \\
 \quad + \alpha_{33}^0\Lambda_{i3}(t_{ij} - \delta) + (\alpha_{31}^0 + \alpha_{31}^1\text{MCI} + \alpha_{31}^2\text{AD})\Lambda_{i1}(t_{ij} - \delta) \\
 \quad + (\alpha_{32}^0 + \alpha_{32}^1\text{MCI} + \alpha_{32}^2\text{AD})\Lambda_{i2}(t_{ij} - \delta) \\
 \quad + (\alpha_{34}^0 + \alpha_{34}^1\text{MCI} + \alpha_{34}^2\text{AD})\Lambda_{i4}(t_{ij} - \delta) \\
 \frac{\Delta\Lambda_{i4}(t_{ij})}{\delta} = \gamma_{04} + \gamma_{14}\text{MCI}_i + \gamma_{24}\text{AD}_i + \gamma_{34}\text{apoe4}_i + v_{i3} \\
 \quad + \alpha_{44}^0\Lambda_{i4}(t_{ij} - \delta) + (\alpha_{41}^0 + \alpha_{41}^1\text{MCI} + \alpha_{41}^2\text{AD})\Lambda_{i1}(t_{ij} - \delta) \\
 \quad + (\alpha_{42}^0 + \alpha_{42}^1\text{MCI} + \alpha_{42}^2\text{AD})\Lambda_{i2}(t_{ij} - \delta) \\
 \quad + (\alpha_{43}^0 + \alpha_{43}^1\text{MCI} + \alpha_{43}^2\text{AD})\Lambda_{i3}(t_{ij} - \delta)
 \end{array} \right. \tag{5.3}$$

avec $\Lambda_{i1}(t_{ij})$, $\Lambda_{i2}(t_{ij})$, $\Lambda_{i3}(t_{ij})$, $\Lambda_{i4}(t_{ij})$ respectivement les dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale », « fonction (Fct) » chez le sujet i au temps t_{ij} , $\Delta\Lambda_{i.} = \Lambda_{i.}(t_{ij}) - \Lambda_{i.}(t_{ij} - \delta)$, et δ le pas de discrétisation ; Les vecteurs β , α et γ incluent les paramètres d'effets fixes (les paramètres β_0 sont fixés 0 par contrainte d'identifiabilité), $(u_{i1}, u_{i2}, u_{i3}, u_{i4}, v_{i1}, v_{i2}, v_{i3}, v_{i4})^\top$ un vecteur d'effets aléatoires centrés de matrice de variance-covariance \mathbf{B} . \mathbf{B} suit la structure indiquée en méthodologie : les intercepts aléatoires sur le niveau initial sont standardisées, les effets aléatoires sont corrélés au sein de chaque dimension latente. Par contre, les effets aléatoires entre dimensions sont indépendants sauf les intercepts aléatoires sur le niveau initial qui captent une potentielle corrélation à l'inclusion.

5.3.3.1 Résultats de la modélisation

Les estimations des paramètres des effets fixes du modèle sont résumés dans le tableau 5.7. Avant d'interpréter les estimations, on peut déjà noter l'amélioration importante du fit du modèle lorsque la dimension cognitive est séparée en deux sous-dimensions. En effet, le modèle considérant 3 dimensions avait un AIC de 38885 points. Celui incluant 4 dimensions (estimé sur exactement les mêmes données) est de 37061 points, soit une amélioration de 1824 points. Cela montre le besoin d'affiner les dimensions cognitives pour bien capter les relations temporelles.

TABLE 5.7 – Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction ».

| L = -18405.63, AIC = 37061.25, p = 125 | | | |
|--|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| Niveau initial | | | |
| β_{11} | -1.014 | 0.101 | < 0.001 |
| β_{21} | -1.690 | 0.128 | < 0.001 |
| β_{31} | -0.529 | 0.062 | < 0.001 |
| β_{41} | -0.352 | 0.080 | < 0.001 |
| β_{51} | -0.115 | 0.105 | 0.274 |
| β_{61} | -0.364 | 0.085 | < 0.001 |
| β_{12} | -1.641 | 0.126 | < 0.001 |
| β_{22} | -2.748 | 0.168 | < 0.001 |
| β_{32} | -0.221 | 0.066 | 0.001 |
| β_{42} | -0.268 | 0.092 | 0.004 |
| β_{52} | 0.285 | 0.120 | 0.018 |
| β_{62} | -0.329 | 0.098 | 0.001 |
| β_{13} | -0.940 | 0.130 | < 0.001 |
| β_{23} | -1.758 | 0.168 | < 0.001 |
| β_{33} | -0.284 | 0.071 | < 0.001 |
| β_{43} | 0.208 | 0.104 | 0.045 |
| β_{53} | 0.504 | 0.136 | < 0.001 |
| β_{63} | -0.139 | 0.108 | 0.197 |
| β_{14} | -1.638 | 0.128 | < 0.001 |
| β_{24} | -3.867 | 0.196 | < 0.001 |
| β_{34} | -0.047 | 0.071 | 0.510 |
| β_{44} | -0.136 | 0.099 | 0.172 |

Suite de la Table 5.7 - Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction ».

| L = -18405.63, AIC = 37061.25, p = 125 | | | |
|--|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| Taux de variation | | | |
| γ_{01} | -0.090 | 0.014 | < 0.001 |
| γ_{11} | -0.004 | 0.018 | 0.824 |
| γ_{21} | -0.117 | 0.052 | 0.024 |
| γ_{31} | 0.036 | 0.010 | < 0.001 |
| γ_{41} | -0.004 | 0.012 | 0.725 |
| γ_{51} | -0.047 | 0.010 | < 0.001 |
| γ_{02} | 0.054 | 0.096 | 0.574 |
| γ_{12} | -0.656 | 0.169 | < 0.001 |
| γ_{22} | -1.933 | 0.377 | < 0.001 |
| γ_{32} | 0.019 | 0.059 | 0.746 |
| γ_{42} | 0.117 | 0.082 | 0.151 |
| γ_{52} | -0.198 | 0.065 | 0.002 |
| γ_{03} | 0.036 | 0.114 | 0.755 |
| γ_{13} | 0.092 | 0.126 | 0.463 |
| γ_{23} | -0.247 | 0.334 | 0.460 |
| γ_{33} | 0.178 | 0.085 | 0.035 |
| γ_{43} | 0.243 | 0.116 | 0.037 |
| γ_{53} | -0.014 | 0.076 | 0.849 |
| γ_{04} | -0.038 | 0.051 | 0.453 |
| γ_{14} | -0.070 | 0.098 | 0.476 |
| γ_{24} | -0.819 | 0.268 | 0.002 |
| γ_{34} | -0.136 | 0.047 | 0.004 |
| Influences temporelles | | | |
| α_{11}^0 | -0.003 | 0.012 | 0.780 |
| α_{12}^0 | 0.016 | 0.009 | 0.069 |
| α_{12}^1 | 0.018 | 0.011 | 0.091 |
| α_{12}^2 | -0.014 | 0.017 | 0.422 |
| α_{13}^0 | -0.006 | 0.008 | 0.419 |
| α_{13}^1 | 0.009 | 0.009 | 0.354 |
| α_{13}^2 | 0.015 | 0.013 | 0.244 |

APPLICATIONS

Suite de la Table 5.7 - Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction ».

| L = -18405.63, AIC = 37061.25, p = 125 | | | |
|--|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| α_{14}^0 | 0.016 | 0.014 | 0.248 |
| α_{14}^1 | -0.007 | 0.014 | 0.634 |
| α_{14}^2 | -0.013 | 0.017 | 0.443 |
| α_{21}^0 | 0.256 | 0.066 | < 0.001 |
| α_{21}^1 | -0.031 | 0.060 | 0.602 |
| α_{21}^2 | -0.217 | 0.075 | 0.004 |
| α_{22}^0 | -0.523 | 0.118 | < 0.001 |
| α_{23}^0 | 0.074 | 0.050 | 0.134 |
| α_{23}^1 | 0.052 | 0.056 | 0.345 |
| α_{23}^2 | 0.036 | 0.072 | 0.616 |
| α_{24}^0 | 0.162 | 0.104 | 0.118 |
| α_{24}^1 | -0.058 | 0.104 | 0.579 |
| α_{24}^2 | -0.170 | 0.117 | 0.145 |
| α_{31}^0 | -0.031 | 0.066 | 0.642 |
| α_{31}^1 | 0.040 | 0.076 | 0.601 |
| α_{31}^2 | -0.221 | 0.092 | 0.017 |
| α_{32}^0 | 0.260 | 0.104 | 0.012 |
| α_{32}^1 | 0.100 | 0.080 | 0.211 |
| α_{32}^2 | 0.094 | 0.114 | 0.408 |
| α_{33}^0 | -0.616 | 0.162 | < 0.001 |
| α_{34}^0 | 0.090 | 0.117 | 0.443 |
| α_{34}^1 | 0.086 | 0.120 | 0.474 |
| α_{34}^2 | 0.146 | 0.137 | 0.287 |
| α_{41}^0 | 0.082 | 0.042 | 0.050 |
| α_{41}^1 | 0.078 | 0.050 | 0.117 |
| α_{41}^2 | -0.132 | 0.063 | 0.036 |
| α_{42}^0 | 0.045 | 0.058 | 0.431 |
| α_{42}^1 | 0.122 | 0.066 | 0.064 |
| α_{42}^2 | 0.030 | 0.090 | 0.741 |
| α_{43}^0 | -0.005 | 0.051 | 0.927 |
| α_{43}^1 | 0.092 | 0.061 | 0.131 |

Suite de la Table 5.7 - Estimations des paramètres du modèle structurel dans la modélisation plus fine des dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et « fonction ».

| L = -18405.63, AIC = 37061.25, p = 125 | | | |
|--|-------------|------------|----------|
| Paramètre | Estimation. | Écart-type | p-valeur |
| α_{43}^2 | 0.177 | 0.075 | 0.018 |
| α_{44}^0 | -0.158 | 0.064 | 0.014 |

Interprétation des paramètres de trajectoires des dimensions

Les résultats précédents pour les dimensions « atrophie de l'hippocampe » et fonctionnelle sont encore retrouvés dans ce modèle bien que l'ajustement sur le niveau cognitif soit plus fin qu'avant. La décomposition de la dimension « cognition » en « mémoire épisodique » et « fluence verbale » apporte des éclairages sur les facteurs de déclin de ces dernières. Considérons notre sujet de référence défini lors de l'interprétation des résultats de la section 5.3.1.1. Un sujet au stade MCI ou AD ou plus âgé ou de sexe masculin ou porteur de l'allèle apolipoprotéine $\epsilon 4$ a un niveau de mémoire épisodique plus bas à l'inclusion tandis qu'un sujet de haut niveau d'éducation à une mémoire épisodique plus élevée à l'inclusion. S'agissant de la fluence verbale, un sujet au stade MCI ou AD ou plus âgé a une fluence verbale plus basse que le sujet de référence à l'inclusion. Par contre, un sujet de sexe masculin ou de haut niveau d'éducation a une fluence verbale plus élevée que le sujet de référence à l'inclusion.

La différence majeure ressortie par cette séparation de la mémoire épisodique et de la fluence verbale est un effet différentiel du sexe sur les fonctions cognitives.

La mémoire épisodique et la fluence verbale n'évoluent pas significativement dans le temps pour le sujet de référence. Par contre, relativement à ce sujet de référence et pour un même état du système, la mémoire épisodique décline plus chez un sujet au stade MCI ou AD ou porteur de l'allèle apolipoprotéine $\epsilon 4$. Au contraire, ces trois variables semblent indépendantes du déclin de la fluence verbale. Seuls le sexe et le niveau d'éducation sont associés à un déclin de la fluence verbale après ajustement sur le niveau du système. En particulier, un sujet de sexe masculin ou de haut niveau d'éducation à un déclin de la fluence verbale ralenti par rapport au sujet de référence. Le sexe masculin et l'éducation seraient donc des facteurs protecteurs du déclin de fluence verbale mais pas de la mémoire épisodique.

Interprétation des paramètres d'influences temporelles

Chez les sujets cognitivement normaux, un volume relatif de l'hippocampe plus faible est

APPLICATIONS

associé à une augmentation du déclin de la mémoire épisodique après ajustement sur les niveaux courants de la fluence verbale et de la fonction.

De même, après ajustement sur la mémoire épisodique et la fluence verbale courantes, un volume relatif de l'hippocampe plus faible semble être associé à une augmentation du déclin de la dimension fonctionnelle. Enfin, après ajustement sur le volume relatif de l'hippocampe courant et sur la dimension fonctionnelle courante, un niveau plus faible de mémoire épisodique est associé à une augmentation du déclin de fluence verbale.

Chez les sujets MCI, et relativement aux sujets normaux, il n'y pas de différence significative de structure à part un lien qui apparaît entre un niveau de mémoire épisodique plus faible et une augmentation du déclin de la dimension fonctionnelle.

Chez les sujets AD, l'influence du volume relatif de l'hippocampe sur la mémoire épisodique et sur le déclin fonctionnel disparaissent. Une association apparaît aussi entre le niveau de fluence verbale et le déclin de fonction.

La figure 5.10 résume les influences temporelles absolues entre les dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et fonction dans les trois stades (CN, MCI, AD) de la maladie d'Alzheimer.

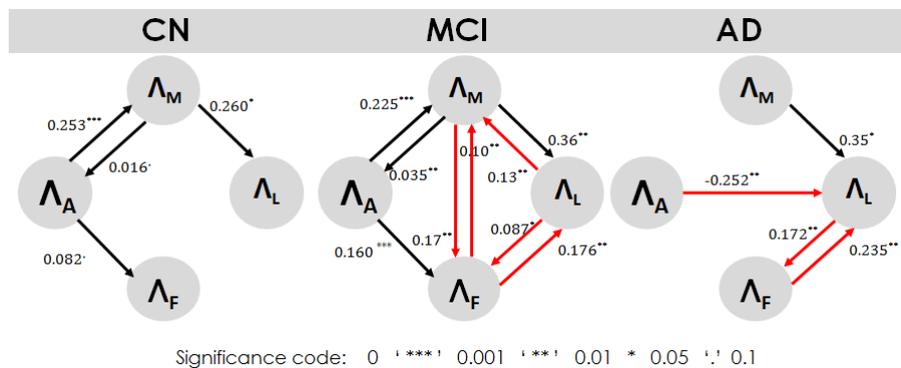


FIGURE 5.10 – Influences temporelles entre les dimensions « atrophie de l'hippocampe » (Λ_A), « mémoire épisodique » (Λ_M), « fluence verbale » (Λ_L) et « fonction » (Λ_F) dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de démence de type Alzheimer).

5.3.3.2 Conclusion

Dans cette deuxième application, nous avons modélisé conjointement de façon plus fine les dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et fonction dans les stades CN (cognitivement normal), MCI (légers déficits cognitifs), AD (diagnostic de

démence de type Alzheimer).

Nous avons adopté une démarche progressive d'exploration plus fine des dimensions anatomique, cognitive et fonctionnelle définies dans la première application. Nous avons construit une dimension plus fine (« atrophie de l'hippocampe ») de la dimension anatomique, deux dimensions « mémoire épisodique » et « fluence verbale » à partir de la dimensions cognitive, la dimension fonctionnelle étant restée inchangée.

Cette définition plus fine des dimensions, nous a permis de mieux comprendre les rôles de l'hippocampe, de la mémoire épisodique et de la fluence verbale dans les trois stades (CN-sujet sans déficits cognitifs, MCI-sujet avec de légers déficits cognitifs, AD-sujet diagnostiqué avec une démence de type Alzheimer) de la maladie d'Alzheimer et des démences apparentées. En particulier, les liens entre hippocampe et cognition passent principalement par la mémoire épisodique, les associations semblant être inexistantes entre hippocampe et fluence verbale ajusté sur le niveau de mémoire épisodique aux stades MCI et CN. Cela peut s'expliquer par les connaissances biologiques, l'hippocampe étant la zone privilégiée de la mémoire.

Tout comme dans la première application, nous avons trouvé des influences temporelles entre les dimensions « atrophie de l'hippocampe », « mémoire épisodique », « fluence verbale » et fonctionnelle dans le stade AD qui pouvaient paraître inattendues (avec des niveaux plus faibles associés à des déclin plus faibles). Cependant, il faut garder en tête que ces associations sont ajustées sur les autres niveaux, il s'agit d'associations résiduelles. Cela peut aussi suggérer un ralentissement de la dynamique chez les patients les plus avancés sur le continuum de la maladie.

Nous avons aussi trouvé dans cette application que le sexe et l'éducation n'influençaient pas de la même manière la mémoire épisodique et la fluence verbale. En effet, le sexe et l'éducation semblent être des facteurs indépendants du déclin de la mémoire épisodique tandis que le sexe masculin et le niveau d'éducation élevé semblent être des facteurs protecteurs de la fluence verbale. Les différences possibles d'effet du sexe et du niveau d'étude en fonction des fonctions cognitives explorées avaient déjà été soulignées dans la littératures [Proust-Lima et al., 2008].

Globalement cette démarche de définition fine des dimensions a permis de mettre en lumière l'importance de définir correctement les dimensions d'intérêt et les marqueurs associés d'une part et d'autre part d'affiner les connaissances sur le mécanisme de la maladie d'Alzheimer notamment entre l'hippocampe, la mémoire épisodique la fluence verbale et la perte d'autonomie du sujet au cours de la maladie.

6 Discussion et Perspectives

Nous avons proposé dans cette thèse une approche statistique innovante pour répondre à deux objectifs fréquents en santé publique lorsqu'on s'intéresse à la progression d'une maladie par le biais de données répétées multivariées. Il s'agit de modéliser conjointement plusieurs marqueurs longitudinaux et comprendre la structure de dépendance entre les marqueurs répétés.

Comme nous l'avons vu dans l'état de l'art, plusieurs approches et techniques existent dans la littérature pour répondre à ces deux objectifs séparément ou simultanément. Toutefois, de nouveaux développements statistiques étaient requis pour atteindre ces objectifs dans le cas de l'étude de maladies complexes comme la maladie d'Alzheimer. Nous avons proposé dans cette thèse une approche statistique originale et l'avons appliquée à la compréhension de la progression des marqueurs dans la maladie d'Alzheimer.

6.1 Approche méthodologique

6.1.1 Objectif

L'objectif statistique de cette thèse était de proposer une approche de modélisation dynamique de plusieurs processus qui permette à la fois de décrire l'évolution des processus et d'explorer leur structure de dépendances afin d'inférer des relations causales entre eux.

6.1.2 Modèle structurel

Par rapport à l'objectif statistique, nous avons proposé une approche méthodologique innovante pour modéliser en population les trajectoires de plusieurs processus latents Gaussiens et leur structures d'influences temporelles. Notre approche définit un modèle structurel à effets mixtes pour les processus latents. Le modèle structurel combine la théorie des modèles linéaires mixtes pour modéliser la dynamique des processus et la théorie des équations de différences pour modéliser les influences temporelles entre les processus latents.

Les équations de différences ont été utilisées par analogie aux équations différentielles pour bénéficier des mêmes interprétations causales des paramètres de ces dernières tout en évitant les problèmes de calcul numérique et d'identifiabilité rencontrés dans l'estimation des modèles à équations différentielles.

Dans notre approche méthodologique, nous avons considéré que les processus latents évoluaient en temps discret avec un pas de discrétisation δ (qui peut être défini suivant les applications).

Nous avons modélisé le niveau initial des processus par un modèle linéaire à effets mixtes. Les corrélations initiales entre les processus sont captés par des effets aléatoires inter-processus corrélés.

La dynamique des processus est modélisée à travers leur taux de variation au cours du temps. Le taux de variation du système de processus entre deux temps consécutifs t et $t + \delta$ est défini par un modèle linéaire à effets mixtes avec une partie auto-régressive d'ordre 1. La partie auto-régressive permet de relier l'état des processus au temps t à leur taux de variation au temps $t + \delta$ grâce à une matrice d'influences temporelles. Cette matrice modélise les corrélations entre les processus sous forme d'influences temporelles.

Les coefficients de la matrice d'influences temporelles peuvent dépendre du temps ou être ajustés sur des variables explicatives pour tenir compte de la dynamique possible des influences entre les processus et de l'effet potentiel de variables explicatives.

Notre approche de modélisation structurelle, en combinant la théorie des modèles linéaires à effets mixtes et la théorie des équations de différences, bénéficie des avantages de ces deux approches de modélisation.

Grâce aux effets aléatoires du modèle, notre approche permet de modéliser plus souplesment la dynamique de plusieurs processus tout en prenant en compte les corrélations entre les valeurs répétées d'un processus chez un même sujet, ainsi que la variabilité inter-sujets du processus.

Grâce à la matrice d'influences, notre approche permet de modéliser à l'échelle de la population les corrélations entre les processus à des fins d'interprétations causales au même titre que les équations différentielles linéaires.

L'approche adoptée pour modéliser les dépendances entre les processus est à mi chemin entre la méthodologie des modèles à équations structurelles (SEM) [Jöreskog, 1970] et celle des modèles à équations différentielles [Commenges et Gégout-Petit, 2009].

Comme dans les modèles à équations structurelles, notre approche sépare bien les observations des quantités latentes d'intérêt. Cependant, par rapport aux modèles à équations structurelles, notre approche a l'avantage de considérer que la quantité sous-jacente modélisée évolue dans le temps même si elle est évaluée à des temps discrets. En effet, dans les approches à équations structurelles, la notion de temps est le plus souvent appréhendée par des variables latentes successives distinctes sans forcément la notion de processus.

Comme nous l'avons vu dans l'état de l'art, des matrices d'influences peuvent être définies dans les deux approches (à équations structurelles et à équations différentielles). Dans notre

approche, la matrice d'influence entre les processus bénéficie d'une interprétation causale plus proche de celle d'un modèle à équations différentielles. En effet, nous ne modélisons pas les influences entre les variables du temps successifs comme dans les modèles SEM mais plutôt l'impact de l'état du système au temps t sur la variation entre deux temps consécutifs t et $t + \delta$.

La limite principale de notre approche par rapport à un modèle à équations différentielles est qu'elle modélise des relations linéaires par intervalle entre les processus. L'approche de modélisation à équation différentielles est en temps continu, plus générale, et des relations non-linéaires peuvent être définies entre les composantes du système de processus. Par contre, cela est souvent fait aux prix des problèmes d'identifiabilité, d'un temps de calcul très long, et de problèmes de résolution numérique d'ODE par un solveur.

Notre approche permet d'éviter ces problèmes d'identifiabilité et de réduire considérablement le temps d'optimisation et la complexité numérique plus généralement.

6.1.3 Modèle d'observation

Pour tenir compte de la nature latente des processus qui sont potentiellement mesurés à travers plusieurs marqueurs, nous avons proposé un modèle d'observation pour les marqueurs continus Gaussiens et non Gaussiens.

Ce modèle d'observation relie un ou plusieurs marqueurs continu à un même processus latent. Nous avons utilisé pour cela une fonction de lien paramétrique linéaire pour les marqueurs Gaussiens et une base de I-splines avec des paramètres positifs pour les marqueurs continus non Gaussiens. Les paramètres du modèle d'observation et du modèle structurel sont estimés simultanément.

En rajoutant un modèle d'observation au modèle structurel, le modèle global n'est pas identifiable. Nous avons contraint judicieusement le niveau initial des processus latents pour fixer leur dimension et modéliser les marqueurs dans leur échelle naturelle.

La prise en compte de marqueurs possiblement non Gaussiens était essentielle dans l'étude de la maladie d'Alzheimer. En effet, nombre de dimensions sont appréhendées par des échelles de mesure aux distributions souvent asymétriques. Il a été montré dans des travaux précédents que ces écarts à la normalité pouvaient induire des biais importants dans les études du vieillissement cognitif [Proust-Lima et al., 2011]. Notre approche de modèle d'observation permet de corriger ces biais potentiels dûs à l'écart à la normalité des marqueurs. Elle offre ainsi une souplesse comparée à d'autres modèles d'observation non paramétrés utilisés dans la majorité des applications. En considérant une famille de fonctions de lien souples et paramétrée, notre approche

permet d'obtenir la transformation la plus vraisemblable pour les données observées.

6.1.4 Package R développé

Notre approche méthodologique a été implémentée dans un package R qui combine la programmation en langages R et C++. Dans sa version actuelle, ce package permet d'estimer un modèle de notre approche de modélisation et d'obtenir les prédictions dans les échelles transformées et naturelles des marqueurs.

Elle permet aussi à l'utilisateur de choisir le pas de discrétisation des processus, les marqueurs de chaque processus et de fixer certains paramètres. Par exemple en fixant les paramètres de la matrice d'influences (la composante auto-régressive) à zéro, on peut estimer un modèle linéaire mixte multivarié à processus latents avec potentiellement plusieurs marqueurs mesurant un même processus latent.

6.1.5 Étude de simulations

Nous avons conduit une étude de simulations pour valider le programme d'estimation et étudier le comportement des estimateurs du maximum de vraisemblance dans le cas d'une mauvaise spécification du modèle.

Les estimateurs du maximum de vraisemblance sont consistants dans le cas d'une bonne spécification de modèle. La structure d'influence temporelles entre les processus est la même que celle obtenue par un modèle à équations différentielles, même si les intensités diffèrent comme montré dans la méthodologie. Ainsi notre approche bénéficie des mêmes interprétations causales que le modèle à équations différentielles linéaires correspondant.

Dans le cas d'une mauvaise spécification des trajectoires des processus et de l'omission de variables explicatives de ces trajectoires, les paramètres peuvent être biaisés et la structure d'influences impactée.

Des précautions doivent donc être prises lors de la spécification des trajectoires des processus et du choix des variables explicatives de ces derniers. Cela peut se faire, comme nous l'avons proposé dans l'application par des études de sensibilité.

6.1.6 Interprétation causale des paramètres de la matrice d'influences temporelles

Nous avons fondamentalement exploré dans cette thèse la dynamique de plusieurs processus et leurs influences temporelles. L'interprétation causale de ces influences doit être faite avec

prudence.

Nous nous sommes appuyés sur l'approche dynamique de la causalité développée notamment par Arjas et Parner [2004], Didelez [2008], Commenges et Gégout-Petit [2009], Aalen et Gunnes [2010]; Aalen et al. [2016]. Comme nous l'avons vu dans l'état de l'art, il existe une vaste littérature statistique sur d'autres approches de causalité, notamment l'approche contrefactuelle Rubin [1974] et l'approche de Granger [Granger, 1969]. Cependant, ces approches ne nous ont pas paru aussi pertinentes au vu de notre objectif. Du fait de la complexité de notre application avec les évolutions parallèles de plusieurs dimensions latentes, toutes potentiellement intéressantes, l'approche contrefactuelle n'était pas adaptée.

L'approche de Granger, par le biais de réseaux bayésiens dynamiques aurait pu être envisagée, et a été initialement envisagée (présentation en conférence lors du 48^{èmes} *Journées de Statistique de la Société Française de Statistique*, Montpellier, France, 2016). Cependant, dans la perspective de comprendre le lien entre les dimensions et de donner plus de considération à l'interprétation causale dans notre approche, nous avons préféré prendre en compte l'aspect dynamique avec une influence des dimensions sur le changement au cours du temps des autres.

Nous avons donc choisi d'analyser les changements des processus (en cherchant des structures de dépendance locales) suivant l'approche dynamique de la causalité [Arjas et Parner, 2004; Didelez, 2008; Commenges et Gégout-Petit, 2009]. Nous avons aussi choisi de définir ces relations de causalité dans un modèle structurel, c'est-à-dire au niveau de processus latents plutôt qu'au niveau des observations. Ce faisant, notre approche se démarque du modèle d'incrément linéaires de Aalen et Gunnes [2010] dont les conclusions causales sont à l'échelle des observations et de ce fait sont très sensibles aux choix des marqueurs des dimensions d'intérêt.

Une des hypothèses majeures de notre approche était la discrétisation du temps. Fondamentalement, les relations causales doivent être explorées à un niveau infinitésimal et par conséquent, un modèle causal doit être défini en temps continu [Commenges et Gégout-Petit, 2009; Aalen et al., 2016]. Notamment Aalen et al. [2016] a récemment montré que l'inférence causale entre processus continus par l'estimation d'un DAG sont biaisés. En effet, l'estimation d'un DAG dans lequel les noeuds sont des mesures transversales de processus potentiellement observés à différents temps introduit des biais considérables sur la structure causale entre les processus. Aalen et al. [2016] ont pointé ainsi l'importance de tenir compte des mesures répétées des processus avec une fréquence d'observation relativement élevée.

Toutefois, la fréquence des observations et la dynamique des processus peuvent être distinguées en spécifiant un modèle d'observation pour des processus évoluant en temps continu [Commenges et Gégout-Petit, 2009] ce qui conduit au modèles à équation différentielles dont les problèmes d'estimation sont mentionnés ci-dessus et qui justifie notre approche en temps discret. Dans notre cas, la discrétisation est un moyen de simplification numérique de l'approche. En particulier, au contraire d'autres approches critiquées, la discrétisation est dissociée de la fré-

quence d'observation. Elle peut donc être faite beaucoup plus finement. Par exemple, pour nos applications, le pas de discrétisation de quelques mois est très faible par rapport à la dynamique de la progression des marqueurs mais aussi par rapport à la fréquence des mesures.

Néanmoins, nous avons souhaité comprendre les liens entre les structures d'influences identifiées dans notre approche et celles obtenues par une approche en temps continu. Nous avons pu établir des relations analytiques après approximation de modèles sous deux pas de discrétisation différents. Nous avons aussi évalué par simulations si notre approche de modélisation bénéficiait des mêmes interprétations causales que les modèles mécanistes et nous avons constaté que l'utilisation d'un modèle en temps discret ne modifiait pas les taux d'erreur de type I des paramètres d'influence temporelle, à condition que le pas de discrétisation reste faible.

Notons enfin que les paramètres d'influences temporelles dans notre approche de modélisation bénéficient d'une interprétation causale sous conditions que l'ensemble des processus définit un système autonome et suffisamment riche [Commenges et Gégout-Petit, 2009] et qu'il n'y pas de facteurs de confusion non mesurés. Ces conditions sont fondamentales pour toute approche d'inférence causale.

6.2 Applications à l'épidémiologie de la maladie d'Alzheimer et des démences apparentées

6.2.1 Objectif

L'objectif épidémiologique de cette thèse était de modéliser conjointement les dimensions anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer et des démences apparentées et d'explorer les relations entre ces dimensions vues comme des processus dynamiques. Cette modélisation vise d'une part à contribuer à une meilleure connaissance de l'histoire naturelle de la maladie d'Alzheimer et d'autre part à valider ou infirmer les schémas théoriques sur la maladie.

6.2.2 Application de notre approche à la maladie d'Alzheimer

Au regard de l'objectif épidémiologique de la thèse, nous avons appliqué notre approche méthodologique sur les données de l'initiative ADNI Weiner et al. [2012] pour modéliser la dynamique des dimensions anatomique, cognitive et fonctionnelle et leur structure d'influences temporelles dans trois stades cliniques de la maladie. Il s'agit des stades CN où le sujet est cognitivement normal - MCI ou le sujet commence à présenter de légers déficits cognitifs objectivés

DISCUSSION ET PERSPECTIVES

par des test psychométriques - AD où le sujet est diagnostiqué en début d'une démence de type Alzheimer.

Dans une première application, nous avons exploré les relations temporelles entre les dimensions anatomique, cognitive et fonctionnelle en supposant des marqueurs très globaux de ces dimensions. Nous avons trouvé des influences temporelles de la dimension anatomique vers les dimensions cognitive et fonctionnelle et de la dimension fonctionnelle vers la dimension cognitive dans le stade CN. Chez les MCI, nous avons aussi trouvé qu'un retour significatif de la dimension cognitive vers les dimensions anatomique et fonctionnelle apparaît. Ces nouvelles relations n'étaient pas forcément retrouvées au stade AD. Cela peut venir de plusieurs raisons. D'abord, nous avons peu de données dans ce stade (avec notamment un suivi plus court) donc les résultats sont à prendre avec précaution. D'autre part la variabilité à ce stade pourrait être moindre et donc des associations plus difficiles à mettre en évidence.

Nous avons vérifié et confirmé la stabilité de ces résultats dans plusieurs analyses de sensibilité étant donné les possibles problèmes de robustesse en cas de mauvaise spécification.

Par la suite, nous avons fait une deuxième application dont l'objectif était de définir plus finement les dimensions anatomique, cognitive et fonctionnelle pour explorer des relations entre des dimensions spécifiques dans la maladie d'Alzheimer et des démences apparentées. Nous avons donc progressivement raffiné les dimensions anatomique et cognitive pour obtenir quatre dimensions fines que sont « l'atrophie de l'hippocampe » mesuré par le volume relatif de l'hippocampe, « la mémoire épisodique » mesurée par le test de fluence verbale (AVLT, pour Auditory Verbal Learning Test) de la catégorie lexicale animale et enfin « la dimension fonctionnelle » mesurée par la FAQ (Functional Assessment Questionnaire).

La modélisation conjointe de ces quatre dimensions a permis de mieux préciser l'effet de variables explicatives sur la dynamique des dimensions. En effet il ressort principalement que l'éducation n'a pas un effet sur le déclin du volume de l'hippocampe et que le sexe masculin et le niveau élevé d'éducation sont des facteurs protecteurs du déclin de la fluence verbale tandis que le sexe et l'éducation sont indépendants du déclin de la mémoire épisodique.

Les influences temporelles entre ces dimensions sont aussi plus claires. Le rôle central de la mémoire épisodique a bien été mis en évidence. C'est cette dimension qui apparaît comme médiatrice dans la relation entre atrophie de l'hippocampe et le déclin de la fluence verbale aux stades « normal » et « déficits cognitifs légers (MCI) ». Ce résultat est particulièrement important pour la compréhension de la maladie. En cela, cette deuxième application nous a permis de mettre en lumière l'importance de définir correctement les dimensions d'intérêt et les marqueurs associés.

6.3 Perspectives

6.3.1 Perspectives de l'approche méthodologique

Les perspectives de l'approche méthodologique sont à court et à long termes.

À court terme, il serait intéressant de :

- proposer des méthodes de validation de l'approche et de vérification de l'adéquation aux données. Par exemple, implémenter des méthodes pour vérifier les hypothèse de normalité des erreurs de mesure du modèle,
- développer un modèle d'observation généralisé pour prendre en compte des données non continues (binaire, catégorielle, ordinale). Cette amélioration fera néanmoins intervenir des difficultés d'estimation puisque la vraisemblance n'aura sans doute plus d'écriture analytique.

À long terme, des développements statistiques sont nécessaires pour :

- prendre en compte des sorties d'étude informatives et des données de survie. En effet l'approche repose actuellement sur l'hypothèse MAR (Missing At Random) qui est raisonnable sur un suivi de courte durée comme c'était le cas dans notre étude sur une durée de 3 ans. Dans un suivi plus long et dans certaines populations comme les cohortes de personnes âgées en population, les sorties d'étude dues par exemple au décès où la démence sont informatives et leur non-prise en compte peut introduit des biais sur les paramètres de notre approche.
- améliorer la modélisation de la structure d'influences temporelles entre les processus. En effet dans notre approche, nous faisons l'hypothèse que les processus sont Markoviens et que conditionnement à l'état du système de processus au temps t l'état du système de processus au temps $t + \delta$ est indépendant des états du processus antérieurs au temps t . Cette hypothèse peut être assouplie en introduisant une composantes auto-régressive d'ordre supérieur à 1.

6.3.2 Perspectives des applications

Quelques perspectives sont aussi à envisager pour améliorer la connaissance de l'histoire naturelle de la maladie d'Alzheimer et des démences apparentées à savoir :

- Répliquer notre analyse sur d'autres cohortes pour valider nos résultats. En effet dans cette thèse, nous nous sommes focalisés sur les données de l'initiative ADNI, mais il peut être intéressant d'appliquer notre modèle sur d'autres jeux de données populationnelles. Des collaboratrices à Montpellier (équipe Inserm U 1061, Neuropsychiatry : Epidemiological and Clinical Research) ont déjà commencé à utiliser l'approche pour décortiquer les

DISCUSSION ET PERSPECTIVES

- relations dynamiques réciproques en déclin cognitif et fonctionnel autour du diagnostic de démence dans la cohorte populationnelle 3 C (article en préparation, Carles et al.)
- Évaluer la dynamique de la structure d'influences entre les dimensions anatomique, cognitive et fonctionnelle. Bien que la méthodologie ait été développée pour inclure possiblement des influences temporelles qui peuvent dépendre du temps, nous nous sommes focalisés dans nos applications sur des relations temporelles indépendantes du temps. Cette perspective est déjà envisagée. Une application est en cours pour étudier les trajectoires de déclin précédant le diagnostic de démence en nous focalisant sur les sujets diagnostiqués déments et en explorant des relations temporelles possiblement dépendantes du temps dans la cohorte populationnelle 3C avec un suivi de 15 ans.
 - Élargir le nombre de dimensions impliquées dans la maladie d'Alzheimer par exemple rajouter plus en amont de la maladie, la dimension biologique avec notamment l'anormalité de la protéine β -Amyloïde. Ce type d'application sera essentiel pour alimenter les discussions en cours sur les biomarqueurs de la maladie d'Alzheimer. Cependant, ces applications sont pour le moment limités par le manque de données répétées sur ces marqueurs pourtant cruciaux.

Bien que proposée et développée pour répondre à des questions dans l'étude de la maladie d'Alzheimer, cette méthodologie est pertinente pour nombre d'autres pathologies qui sont caractérisées par des évolutions progressives de plusieurs dimensions. Elle devrait donc trouver des applications beaucoup plus larges.

Enfin, au delà de la compréhension des mécanismes entre des marqueurs de progression, cette méthodologie sera aussi potentiellement très utile pour décrire des facteurs de risque (ou expositions) dépendantes du temps sur le développement de maladies.

Bibliographie

- Aalen O. O. et Gunnes N. A dynamic approach for reconstructing missing longitudinal data using the linear increments model. *Biostatistics*, 11(3) : 453–472, 2010. 55, 155
- Aalen O. O., Røysland K., Gran J. M., Kouyos R., et Lange T. Can we believe the DAGs? A comment on the relationship between causal DAGs and mechanisms. *Statistical methods in medical research*, 25(5) : 2294–2314, 2016. 53, 55, 155
- Akaike H. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory, (1973)*, pages 267–281, 1973. 34
- Almeida O. P., Hulse G. K., Lawrence D., et Flicker L. Smoking as a risk factor for Alzheimer’s disease : Contrasting evidence from a systematic review of case–control and cohort studies. *Addiction*, 97(1) : 15–28, 2002. 22
- Amieva H., Le Goff M., Millet X., Orgogozo J. M., Pérès K., Barberger-Gateau P., Jacqmin-Gadda H., et Dartigues J. F. Prodromal Alzheimer’s disease : Successive emergence of the clinical symptoms. *Annals of Neurology*, 64(5) : 492–498, 2008. 21
- Arjas E. et Parner J. Causal reasoning from longitudinal data. *Scandinavian Journal of Statistics*, 31(2) : 171–187, 2004. 42, 53, 63, 155
- Asparouhov T., Hamaker E. L., et Muthén B. Dynamic Structural Equation Models. *Structural Equation Modeling : A Multidisciplinary Journal*, 25(3) : 359–388, 2017. 61, 62
- Baker F. B. *The basics of item response theory*. ERIC, 2001. 35
- Bollen K. A. et Pearl J. Eight myths about causality and Structural Equation Models. In *Handbook of causal analysis for social research*, pages 301–328. Springer, 2013. 40
- Broe G. A., Grayson D. A., Creasey H. M., Waite L. M., Casey B. J., Bennett H. P., Brooks W. S., et Halliday G. M. Anti-inflammatory drugs protect against Alzheimer disease at low doses. *Archives of Neurology*, 57(11) : 1586–1591, 2000. 22
- Brookmeyer R., Johnson E., Ziegler-Graham K., et Arrighi H. M. Forecasting the global burden of Alzheimer’s disease. *Alzheimer’s & dementia*, 3(3) : 186–191, 2007. 20
- Byrne B. M. *Structural Equation Modeling with Mplus : Basic concepts, applications, and programming*. Routledge, 2013. 61
- Caramelli P., Carthery-Goulart M. T., Porto C. S., Charchat-Fichman H., et Nitrini R. Category fluency as a screening test for Alzheimer disease in illiterate and literate patients. *Alzheimer Disease & Associated Disorders*, 21(1) : 65–67, 2007. 138
- Cheung M. W.-L. *Meta-analysis : A Structural Equation Modeling approach*. John Wiley & Sons, 2015. 58
- Commenges D. et Gégout-Petit A. A general dynamical statistical model with causal interpretation. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 71(3) : 719–736, 2009. 41, 63, 65, 66, 152, 155, 156

BIBLIOGRAPHIE

- Crane P. K., Carle A., Gibbons L. E., Insel P., Mackin R. S., Gross A., Jones R. N., Mukherjee S., Curtis S. M., Harvey D., et al. Development and assessment of a composite score for memory in the Alzheimer's Disease Neuroimaging Initiative (ADNI). *Brain imaging and behavior*, 6(4) : 502–516, 2012. 118
- Dempster A. P., Laird N. M., et Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38, 1977. 32
- Desmée S., Mentré F., Veyrat-Follet C., Sébastien B., et Guedj J. Using the SAEM algorithm for mechanistic joint models characterizing the relationship between nonlinear psa kinetics and survival in prostate cancer patients. *Biometrics*, 73(1) : 305–312, 2017. 69
- Dickerson B. C., Bakkour A., Salat D. H., Feczko E., Pacheco J., Greve D. N., Grodstein F., Wright C. I., Blacker D., Rosas H. D., et al. The cortical signature of Alzheimer's disease : Regionally specific cortical thinning relates to symptom severity in very mild to mild AD dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral cortex*, 19(3) : 497–510, 2008. 118
- Didelez V. Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 70(1) : 245–264, 2008. 41, 43, 155
- Diggle P., Farewell D., et Henderson R. Analysis of longitudinal data with drop-out : Objectives, assumptions and a proposal. *Journal of the royal statistical society : Series C (Applied Statistics)*, 56(5) : 499–550, 2007. 55
- Diggle P. J. An approach to the analysis of repeated measurements. *Biometrics*, pages 959–971, 1988. 35
- Duncan O. D. *Introduction to Structural Equation models*. Elsevier, 1975. 27, 58
- Dunson D. B. Dynamic latent trait models for multidimensional longitudinal data. *Journal of the American Statistical Association*, 98(463) : 555–563, 2003. 33, 60, 62
- Dureman I. et Sälde H. Psychometric and experimental methods for the clinical evaluation of mental functioning. *Uppsala : Almqvist & Wiksell*, 1959. 27
- Dutcher S. K., Rattinger G. B., Langenberg P., Chhabra P. T., Liu X., Rosenberg P. B., Leoutsakos J.-M., Simoni-Wastila L., Walker L. D., Franey C. S., et al. Effect of medications on physical function and cognition in nursing home residents with dementia. *Journal of the American Geriatrics Society*, 62(6) : 1046–1055, 2014. 49
- Eddelbuettel D. et Sanderson C. Rcpparmadillo : Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis*, 71 : 1054–1063, 2014. 96
- Ehrenstein G., Galdzicki Z., et Lange G. D. The choline-leakage hypothesis for the loss of acetylcholine in Alzheimer's disease. *Biophysical journal*, 73(3) : 1276, 1997. 67, 69
- Fieuws S. et Verbeke G. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2) : 424–431, 2006. 37, 38
- Fleminger S., Oliver D., Lovestone S., Rabe-Hesketh S., et Giora A. Head injury as a risk factor for Alzheimer's disease : The evidence 10 years on ; a partial replication. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(7) : 857–862, 2003. 22
- Florens J.-P. et Fougere D. Noncausality in continuous time. *Econometrica : Journal of the Econometric Society*, pages 1195–1212, 1996. 66
- Folstein M. F., Folstein S. E., et McHugh P. R. “mini-mental state” : A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3) : 189–198, 1975. 116

- Friedman N., Geiger D., et Goldszmidt M. Bayesian network classifiers. *Machine learning*, 29 (2-3) : 131–163, 1997. 52
- Gao S., Hendrie H. C., Hall K. S., et Hui S. The relationships between age, sex, and the incidence of dementia and Alzheimer disease : A meta-analysis. *Archives of general psychiatry*, 55(9) : 809–815, 1998. 21
- Ghahramani Z. Learning dynamic Bayesian networks. In *Adaptive processing of sequences and data structures*, pages 168–197. Springer, 1998. 53
- Ghahramani Z. et Hinton G. E. Parameter estimation for linear dynamical systems. Technical report, Technical Report CRG-TR-96-2, University of Toronto, Dept. of Computer Science, 1996. 53
- Gomez R. G. et White D. A. Using verbal fluency to detect very mild dementia of the Alzheimer type. *Archives of Clinical Neuropsychology*, 21(8) : 771–775, 2006. 138
- Grady C. L., McIntosh A. R., et Craik F. I. Age-related differences in the functional connectivity of the hippocampus during memory encoding. *Hippocampus*, 13(5) : 572–586, 2003. 118
- Granger C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica : Journal of the Econometric Society*, pages 424–438, 1969. 44, 49, 63, 155
- Green P. J. et Silverman B. W. *Nonparametric regression and generalized linear models : A roughness penalty approach*. CRC Press, 1993. 33
- Greenland S. Quantifying biases in causal models : Classical confounding vs collider-stratification bias. *Epidemiology*, pages 300–306, 2003. 63
- Guedj J., Thiébaud R., et Commenges D. Maximum likelihood estimation in dynamical models of HIV. *Biometrics*, 63(4) : 1198–1206, 2007. 68
- Hall C. B., Ying J., Kuo L., Sliwinski M., Buschke H., Katz M., et Lipton R. B. Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in medicine*, 20(24) : 3695–3714, 2001. 38
- Han S. D., Gruhl J., Beckett L., Dodge H. H., Stricker N. H., Farias S., Mungas D., Alzheimer’s Disease Neuroimaging Initiative, et al. Beta amyloid, tau, neuroimaging, and cognition : Sequence modeling of biomarkers for Alzheimer’s disease. *Brain imaging and behavior*, 6(4) : 610–620, 2012. 28
- Hernán M. A., Brumback B. A., et Robins J. M. Estimating the causal effect of zidovudine on CD4 count with a Marginal Structural Model for repeated measures. *Statistics in medicine*, 21(12) : 1689–1709, 2002. 46, 48, 49
- Hernán M. A., Hernández-Díaz S., et Robins J. M. A structural approach to selection bias. *Epidemiology*, pages 615–625, 2004. 63
- Holland P. W., Glymour C., et Granger C. Statistics and causal inference. *ETS Research Report Series*, 1985(2), 1985. 44, 45, 46
- Jack C. R., Knopman D. S., Jagust W. J., Shaw L. M., Aisen P. S., Weiner M. W., Petersen R. C., et Trojanowski J. Q. Hypothetical model of dynamic biomarkers of the Alzheimer’s pathological cascade. *The Lancet Neurology*, 9(1) : 119–128, 2010. 17, 22, 23
- Jack C. R., Knopman D. S., Jagust W. J., Petersen R. C., Weiner M. W., Aisen P. S., Shaw L. M., Vemuri P., Wiste H. J., Weigand S. D., et al. Tracking pathophysiological processes in Alzheimer’s disease : An updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12(2) : 207–216, 2013. 23

BIBLIOGRAPHIE

- Jack C. R., Bennett D. A., Blennow K., Carrillo M. C., Dunn B., Haeblerlein S. B., Holtzman D. M., Jagust W., Jessen F., Karlawish J., et al. NIA-AA Research Framework : Toward a biological definition of Alzheimer's disease. *Alzheimer's & Dementia*, 14(4) : 535–562, 2018. 21, 23
- Jacqmin-Gadda H., Commenges D., et Dartigues J.-F. Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics*, 62(1) : 254–260, 2006. 33
- Joly P. *Estimation de la fonction de risque dans un contexte général de troncature et de censure : Application à l'estimation de l'incidence de la démence*. PhD thesis, Bordeaux 2, 1996. 77
- Joly P., Commenges D., Helmer C., et Letenneur L. A penalized likelihood approach for an illness–death model with interval-censored data : Application to age-specific incidence of dementia. *Biostatistics*, 3(3) : 433–443, 2002. 22
- Jöreskog K. G. A general method for estimating a linear Structural Equation System. *ETS Research Bulletin Series*, 1970(2) : i–41, 1970. 60, 152
- Jöreskog K. G. Analysis of covariance structures. In *Multivariate Analysis–III*, pages 263–285. Elsevier, 1973. 58
- Jorm A. F. History of depression as a risk factor for dementia : An updated review. *Australian & New Zealand Journal of Psychiatry*, 35(6) : 776–781, 2001. 22
- Jorm A. F. et Jolley D. The incidence of dementia a meta-analysis. *Neurology*, 51(3) : 728–733, 1998. 21, 22
- Jorm A. F., Creasey H., Broe G. A., Sulway M. R., Kos S. C., et Dent O. F. The advantage of being broad-minded : Brain diameter and neuropsychological test performance in elderly war veterans. *Personality and individual differences*, 23(3) : 371–377, 1997. 22
- Kjærulff U. dhugin : A computational system for dynamic time-sliced Bayesian networks. *International journal of forecasting*, 11(1) : 89–111, 1995. 53
- Kohs S. C. Intelligence measurement : A psychological and statistical study based upon the block-design tests. 1923. 38
- Korb K. B. et Nicholson A. E. *Bayesian artificial intelligence*. CRC press, 2010. 54
- Laird N. M. Missing data in longitudinal studies. *Statistics in medicine*, 7(1-2) : 305–315, 1988. 33
- Laird N. M. et Ware J. H. Random-effects models for longitudinal data. *Biometrics*, pages 963–974, 1982. 26, 31
- Landau S. M., Harvey D., Madison C. M., Koeppe R. A., Reiman E. M., Foster N. L., Weiner M. W., Jagust W. J., Alzheimer's Disease Neuroimaging Initiative, et al. Associations between cognitive, functional, and FDG-PET measures of decline in AD and MCI. *Neurobiology of aging*, 32(7) : 1207–1218, 2011. 28
- Lange N., Carlin B. P., et Gelfand A. E. Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *Journal of the American Statistical Association*, 87(419) : 615–626, 1992. 32, 33
- Lauritzen S. L. et Wermuth N. Graphical models for associations between variables, some of which are qualitative and some quantitative. *The annals of Statistics*, pages 31–57, 1989. 41
- Letenneur L., Gilleron V., Commenges D., Helmer C., Orgogozo J.-M., et Dartigues J.-F. Are sex and educational level independent predictors of dementia and Alzheimer's disease ? Incidence data from the PAQUID project. *Journal of Neurology, Neurosurgery & Psychiatry*, 66(2) : 177–183, 1999. 22

- Liang H. et Wu H. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103(484) : 1570–1583, 2008. 68
- Lindstrom M. J. et Bates D. M. Newton–Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404) : 1014–1022, 1988. 32
- Marquardt D. W. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2) : 431–441, 1963. 32, 85, 86, 96
- McLaughlin T., Buxton M., Mittendorf T., Redekop W., Mucha L., Darba J., Jonsson L., Lacey L., et Leibman C. Assessment of potential measures in models of progression in Alzheimer disease. *Neurology*, 75(14) : 1256–1262, 2010. 17, 23, 25
- Monolix. Version 2018r1 Antony, France : Lixoft SAS, 2018. <http://lixoft.com/products/monolix/>. 68
- Mormino E., Kluth J., Madison C., Rabinovici G., Baker S., Miller B., Koeppe R., Mathis C., Weiner M., Jagust W., et al. Episodic memory loss is related to hippocampal-mediated β -amyloid deposition in elderly subjects. *Brain*, 132(5) : 1310–1323, 2008. 133
- Morris J. C. The clinical dementia rating (cdr) : Current version and scoring rules. *Neurology*, 1993. 116
- Mueller S. G., Weiner M. W., Thal L. J., Petersen R. C., Jack C., Jagust W., Trojanowski J. Q., Toga A. W., et Beckett L. The Alzheimer’s Disease Neuroimaging Initiative. *Neuroimaging Clinics of North America*, 15(4) : 869–877, 2005. 117
- Mungas D., Harvey D., Reed B., Jagust W., DeCarli C., Beckett L., Mack W., Kramer J., Weiner M., Schuff N., et al. Longitudinal volumetric MRI change and rate of cognitive decline. *Neurology*, 65(4) : 565–571, 2005. 38
- Murphy K. P. et Russell S. Dynamic Bayesian networks : Representation, inference and learning. 2002. 53
- Muthén B. Latent variable modeling of longitudinal and multilevel data. *Sociological methodology*, 27(1) : 453–480, 1997. 60, 62
- Mykland A. P. Statistical causality. *University of Bergen*, 1986. 44, 63, 64, 66
- Orgogozo J., Dartigues J., Lafont S., Letenneur L., Commenges D., Salamon R., Renaud S., et Breteler M. Wine consumption and dementia in the elderly : A prospective community study in the Bordeaux area. *Revue neurologique*, 153(3) : 185–192, 1997. 22
- Pearl J. Causality : Models, reasoning, and inference. *Econometric Theory*, 19(675–685) : 46, 2003. 51, 62
- Pearl J. *Causality*. Cambridge university press, 2009. 42
- Pearl J. Causal diagrams for empirical research. *Biometrika*, 82(4) : 669–688, 1995. 63
- Petrović L. Causality and Markovian representations. *Statistics & probability letters*, 29(3) : 223–227, 1996. 63, 65, 66
- Pfeffer R. I., Kurosaki T. T., Harrah C. H., Chance J. M., et Filos S. Measurement of functional activities in older adults in the community. *Journal of Gerontology*, 37(3) : 323–329, 1982. 118
- Prague M., Commenges D., Guedj J., Drylewicz J., et Thiébaud R. NIMROD : A program for inference via a normal approximation of the posterior in models with random effects based on ordinary differential equations. *Computer methods and programs in biomedicine*, 111(2) : 447–458, 2013. 68

BIBLIOGRAPHIE

- Prague M., Commenges D., Gran J. M., Ledergerber B., Young J., Furrer H., et Thiébaud R. Dynamic models for estimating the effect of HAART on CD4 in observational studies : Application to the Aquitaine Cohort and the Swiss HIV Cohort Study. *Biometrics*, 73(1) : 294–304, 2016. 68
- Proust-Lima C., Jacqmin-Gadda H., Taylor J. M., Ganiayre J., et Commenges D. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, 62(4) : 1014–1024, 2006. 27, 34, 35, 76
- Proust-Lima C., Amieva H., Dartigues J.-F., et Jacqmin-Gadda H. Sensitivity of four psychometric tests to measure cognitive changes in brain aging-population-based studies. *American journal of epidemiology*, 165(3) : 344–350, 2007. 35
- Proust-Lima C., Amieva H., Letenneur L., Orgogozo J.-M., Jacqmin-Gadda H., et Dartigues J.-F. Gender and education impact on brain aging : A general cognitive factor approach. *Psychology and aging*, 23(3) : 608, 2008. 150
- Proust-Lima C., Dartigues J.-F., et Jacqmin-Gadda H. Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *American journal of epidemiology*, 174(9) : 1077–1088, 2011. 26, 153
- Proust-Lima C., Amieva H., et Jacqmin-Gadda H. Analysis of multivariate mixed longitudinal data : A flexible latent process approach. *British Journal of Mathematical and Statistical Psychology*, 66(3) : 470–487, 2013. 34, 35, 76
- Ramsay J. O. Monotone regression splines in action. *Statistical Science*, 3(4) : 425–441, 1988. 77
- Ramsay J. O., Hooker G., Campbell D., et Cao J. Parameter estimation for differential equations : A generalized smoothing approach. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 69(5) : 741–796, 2007. 68
- Reitz C. et Mayeux R. Alzheimer disease : Epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology*, 88(4) : 640–651, 2014. 21
- Rey A. L'examen clinique en psychologie. *Paris : Presse Universitaire de France*, pages 963–974, 1964. 138
- Ricci M., Graef S., Blundo C., et Miller L. A. Using the rey auditory verbal learning test (RAVLT) to differentiate Alzheimer's dementia and behavioural variant fronto-temporal dementia. *The Clinical Neuropsychologist*, 26(6) : 926–941, 2012. 138
- Robins J. M., Hernan M. A., et Brumback B. Marginal Structural Models and causal inference in epidemiology. *Epidemiology*, 11(5) : 550–560, 2000. 46, 47, 48, 49
- Robitaille A., Muniz G., Piccinin A. M., Johansson B., et Hofer S. M. Multivariate longitudinal modeling of cognitive aging : Associations among change and variation in processing speed and visuospatial ability. *GeroPsych : The Journal of Gerontopsychology and Geriatric Psychiatry*, 25(1) : 15, 2012. 27, 38, 39
- Rosenbaum P. R. et Rubin D. B. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1) : 41–55, 1983. 45, 46, 47
- Rubin D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology*, 66(5) : 688, 1974. 44, 155
- Sánchez B. N., Budtz-Jørgensen E., Ryan L. M., et Hu H. Structural equation models : a review with applications to environmental epidemiology. *Journal of the American Statistical Association*, 100(472) : 1443–1455, 2005. 83
- Shrier I. et Platt R. W. Reducing bias through Directed Acyclic Graphs. *BMC medical research methodology*, 8(1) : 70, 2008. 63

- Soetaert K., Petzoldt T., et Setzer R. W. Solving differential equations in R : Package desolve. *Journal of Statistical Software*, 33(9) : 1–25, 2010. 113
- Song L., Kolar M., et Xing E. P. Time-Varying Dynamic Bayesian Networks. pages 1732–1740, 2009. 54
- Sperling R. A., Aisen P. S., Beckett L. A., Bennett D. A., Craft S., Fagan A. M., Iwatsubo T., Jack Jr C. R., Kaye J., Montine T. J., et al. Toward defining the preclinical stages of Alzheimer’s disease : Recommendations from the National Institute on Aging-Alzheimer’s association workgroups on diagnostic guidelines for Alzheimer’s disease. *Alzheimer’s & dementia*, 7 (3) : 280–292, 2011. 17, 23, 24
- Sy J., Taylor J., et Cumberland W. A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics*, pages 542–555, 1997. 37, 38
- Thiébaud R., Jacqmin-Gadda H., Chêne G., Leport C., et Commenges D. Bivariate linear mixed models using SAS proc MIXED. *Computer methods and programs in biomedicine*, 69 (3) : 249–256, 2002. 38
- Verbeke G. Linear mixed models for longitudinal data. In *Linear mixed models in practice*, pages 63–153. Springer, 1997. 32
- Verbeke G. et Lesaffre E. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics & Data Analysis*, 23(4) : 541–556, 1997. 33
- Verbyla A. P., Cullis B. R., Kenward M. G., et Welham S. J. The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society : Series C (Applied Statistics)*, 48(3) : 269–311, 1999. 33
- Wang Y., Chen K., Zhang J., Yao L., Li K., Jin Z., Ye Q., et Guo X. Aging influence on gray matter structural associations within the default mode network utilizing Bayesian network modeling. *Frontiers in aging neuroscience*, 6 : 105, 2014. 54
- Warrington N. M., Tilling K., Howe L. D., Paternoster L., Pennell C. E., Wu Y. Y., et Briollais L. Robustness of the linear mixed effects model to error distribution assumptions and the consequences for genome-wide association studies. *Statistical applications in genetics and molecular biology*, 13(5) : 567–587, 2014. 33
- Wechsler D. *WISC-III : Wechsler intelligence scale for children : Manual*. Psychological Corporation, 1991. 27, 38
- Weiner M. W., Veitch D. P., Aisen P. S., Beckett L. A., Cairns N. J., Green R. C., Harvey D., Jack C. R., Jagust W., Liu E., et al. The Alzheimer’s Disease neuroimaging Initiative : A review of papers published since its inception. *Alzheimer’s & Dementia*, 9(5) : e111–e194, 2012. 20, 30, 156
- Wermuth N. et Lauritzen S. L. On substantive research hypotheses, conditional independence graphs and graphical chain models. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 21–50, 1990. 40
- Wickett J. C., Vernon P. A., et Lee D. H. Relationships between factors of intelligence and brain volume. *Personality and Individual Differences*, 29(6) : 1095–1122, 2000. 22
- Wiener N. The theory of prediction. *Modern mathematics for engineers*, 1956. 49
- Wilkosz P. A., Seltman H. J., Devlin B., Weamer E. A., Lopez O. L., DeKosky S. T., et Sweet R. A. Trajectories of cognitive decline in Alzheimer’s disease. *International Psychogeriatrics*, 22(2) : 281–290, 2010. 33
- Zeger S. L. et Liang K.-Y. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, pages 121–130, 1986. 33

Annexe

6.1 Article méthodologique

DYNAMIC MODELING OF MULTIVARIATE LATENT PROCESSES AND THEIR TEMPORAL RELATIONSHIPS: APPLICATION TO ALZHEIMER'S DISEASE

Bachirou O. Taddé^{1,*}, H el ene Jacqmin-Gadda¹, Jean-Fran ois Dartigues¹, Daniel Commenges¹, C ecile Proust-Lima¹ for the Alzheimer's Disease Neuroimaging Initiative[†]

¹ INSERM, UMR1219, Univ. Bordeaux, F33000 Bordeaux, France

* *oladedji-bachirou.tadde@u-bordeaux.fr*

[†]*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/show_to_apply/ADNI_Acknowledgement_List.pdf*

Abstract

Alzheimer's disease gradually affects several components including the cerebral dimension with brain atrophies, the cognitive dimension with a decline in various functions and the functional dimension with impairment in the daily living activities. Understanding how such dimensions interconnect is crucial for AD research. However it requires to simultaneously capture the dynamic and multidimensional aspects, and to explore temporal relationships between dimensions. We propose an original dynamic model that accounts for all these features. The model defines dimensions as latent processes and combines a multivariate linear mixed model and a system of difference equations to model trajectories and temporal relationships between latent processes in finely discrete time. Parameters are estimated in the maximum likelihood framework enjoying a closed form for the likelihood. We demonstrate in a simulation study that this dynamic model in discrete time benefits the same causal interpretation of temporal relationships as mechanistic models defined in continuous time. The model is then applied to the data of the Alzheimer's Disease Neuroimaging Initiative. Three longitudinal dimensions (cerebral anatomy, cognitive ability and functional autonomy) are analyzed and their temporal structure is contrasted between different clinical stages of Alzheimer's disease.

Keywords. causality, mixed models, difference equations, latent process, longitudinal data.

1 Introduction

Dementia is a general syndrome characterized by a long term and gradual decrease in the ability to think and remember with consequences on the person’s daily functioning. It represents a pressing public health problem and major research priority. Alzheimer disease (AD) is the most common form of dementia, representing 60% to 80% of the cases (Reitz and Mayeux, 2014). AD gradually affects multiple components long before any clinical diagnosis including brain atrophies, cognitive decline in various functions (memory, language, orientation in space and time, etc.) and loss of autonomy in the daily living activities. There is currently no cure for AD although many novel compounds are under development. Natural history of AD and its progression are still misunderstood; yet their understanding constitutes a crucial step for the assessment of compounds, and more generally AD research. Theoretical schemes have been proposed to understand the disease taken as a whole, highlighting the expected dynamic and multidimensional aspects (Jack et al., 2013). Jack’s scheme, which is now central in AD research, hypothesizes a sequence of alterations: mainly the accumulation of proteins in the brain (amyloid- β and tau proteins), the atrophy of brain regions (e.g., hippocampus) and finally clinical manifestations with cognitive and functional declines on which the diagnosis of dementia currently relies. However, because of its complexity, such theoretical scheme is very difficult to translate into a statistical model as it necessitates to combine the dynamic and multidimensional aspects, and to properly explore the temporal relationships between dimensions.

The dynamic aspect of AD has been apprehended mostly using the linear mixed model theory with models that analyzed one dimension according to time and covariates (Hall et al., 2000; Amieva et al., 2008; Donohue et al., 2014). One difficulty is that the dimension of interest is usually not directly observed but measured by a series of outcomes; for instance, cognition is usually measured by a battery of neuropsychological tests. Proust-Lima et al. (2006, 2013) considered latent process models which distinguish a structural model for analyzing the unobserved quantity over time and equations of observations for linking the unobserved quantity to the observed outcomes.

To account for the multidimensional aspect of AD and understand how dimensions are inter-related, many contributions explored pre-determined relationships by examining change over time of one biomarker (e.g., cognitive measure) according to another observed biomarker (e.g., hippocampal initial volume or change) and assumed the latter was ob-

served without measurement error (e.g. Landau et al. (2011); Han et al. (2012)). This approach quantifies temporal relationships but it relies on a specific a priori determined sequence and does not consider all biomarkers as stochastic processes, which limits the interpretation and may induce biases. Some authors also proposed to use bivariate mixed models to dynamically model two dimensions and account for their correlation through correlated random effects (e.g., (Mungas et al., 2005; Robitaille et al., 2012)). However such models remain descriptive; they do not rely on asymmetric temporal relationships between processes and thus do not allow causal interpretations.

Temporal asymmetric relationships between processes have usually been apprehended with Dynamic Bayesian Networks (DBN) (Song et al., 2009). This approach extends the concept of Directed Acyclic Graphs (DAG) (Greenland, 2000) to the dynamic case by modelling constant or time-varying temporal relationships between successive states of a network of processes. However, firstly time is usually too grossly discretized so that spurious temporal associations might appear as showed by Aalen et al. (2016). Second, the trajectories of processes are not modelled over time. Finally, DBN quantify the association between successive levels of processes while a dynamic view of causality would seek local dependence structures linking the network of processes to its infinitesimal change over time (Aalen and Frigessi, 2007).

Local dependence structures can be naturally investigated with mechanistic models which are dynamic models relating a system of processes over time using differential equations. Proposed in HIV studies (Prague et al., 2017) or cancer studies (Desmée et al., 2017), they allowed retrieving causal associations between disease components. Yet mechanistic models are numerically very demanding so that their application to complex diseases such as AD is compromised. In addition they require precise biological knowledge which lacks for AD.

The objective of this work is to propose a statistical model that simultaneously describes the dynamics of multiple dimensions involved in Alzheimer’s disease and assess their temporal relationships similarly as in a mechanistic model but with much less numerical complexity. We consider a system of latent processes, each one representing a latent dimension possibly observed through one or several longitudinal markers. In contrast with mechanistic models, we rely on a discrete-time framework and define a set of difference equations to model the change over time of the system according to the previous state of the system. In addition, we account for individual differences through subject-specific covariates and random effects. As discretization can distort the causal interpretations of

temporal relationships compared to a model in continuous time, we specifically evaluate the impact of discretization on the temporal influence structure in a simulation study. The methodology is applied to the Alzheimer’s Disease Neuroimaging Initiative database to explore the temporal structure between cerebral, cognitive and functional dimensions at different stages of AD.

2 Motivating Data

Data used in the preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD). For up-to-date information, see www.adni-info.org. We focused in this work on the ADNI 1 phase which is a longitudinal multisite observational study that included approximately 800 individuals aged between 55-90 enrolled in three different stages of progression to AD: normal aging (CN, $N \approx 200$), mild cognitive impairment (MCI, $N \approx 400$), and diagnosed Alzheimer’s Disease (dAD, $N \approx 200$). The individuals were followed-up every 6 months up to 3 years for CN and MCI groups and up to 2 years for the dAD group. At each visit, a lot of information was collected, including for the present work, volumes of brain regions measured by MRI, a battery of 19 cognitive tests and a functional assessment with the Functional Activities Questionnaire(FAQ) (Pfeffer et al., 1982). Further details on the ADNI 1 study can be found in Mueller et al. (2005).

3 Methodology

3.1 Structural model for the system of latent processes

Consider D latent processes $\mathbf{\Lambda}_i(t)_{t \geq 0}$ (with $\mathbf{\Lambda}_i(t) = (\Lambda_i^d(t))_{d=1, \dots, D}^\top$) representing a system of D dimensions (e.g. cerebral anatomy, cognitive ability and functional autonomy dimensions for Alzheimer’s disease) for individual i with $i = 1, \dots, N$. We assume $\mathbf{\Lambda}_i(t)_{t \geq 0}$ is defined at discrete times $t = j \times \delta$ with $j = 0, 1, \dots, J$, and δ a constant discretization

step. Let us denote $\Delta\mathbf{\Lambda}_i(t + \delta) = \mathbf{\Lambda}_i(t + \delta) - \mathbf{\Lambda}_i(t)$ the change of the system between two successive times and $\frac{\Delta\mathbf{\Lambda}_i(t+\delta)}{\delta}$ the rate of change of the system. The model for the trajectories of the processes is split in two parts: (i) the level of the processes at baseline $\mathbf{\Lambda}_i(0)$ is modelled with a multivariate linear mixed model, and (ii) the rate of change of the system over time $\frac{\Delta\mathbf{\Lambda}_i(t+\delta)}{\delta}$ is modelled by difference equations combined with a multivariate linear mixed model:

$$\begin{cases} \mathbf{\Lambda}_i(0) = \mathbf{X}_i^0\boldsymbol{\beta} + \mathbf{u}_i \\ \frac{\Delta\mathbf{\Lambda}_i(t+\delta)}{\delta} = \mathbf{X}_i(t + \delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta)\mathbf{v}_i + \mathbf{A}_{i,\delta}(t)\mathbf{\Lambda}_i(t), \quad \forall t > 0, \end{cases} \quad (1)$$

where \mathbf{X}_i^0 is the $D \times p_0$ -matrix of covariates associated with the p_0 -vector of fixed effects $\boldsymbol{\beta}$, and \mathbf{u}_i is the D -vector of individual random intercepts u_i^d in the initial system $\mathbf{\Lambda}_i(0)$. The $(D \times p)$ -matrix \mathbf{X}_i and the $(D \times q)$ -matrix \mathbf{Z}_i include time-dependent covariates associated with the p -vector of fixed effects $\boldsymbol{\gamma}$ and the q -vector ($q = \sum_{d=1}^D q_d$) of individual random effects $\mathbf{v}_i = (v_i^{d\top})_{d=1,\dots,D}$, respectively. $\mathbf{A}_{i,\delta}(t)$ is the $D \times D$ -matrix of transition intensities.

For each dimension d , the $(q_d + 1)$ vector of individual random effects $(u_i^d, v_i^{d\top})^\top$ is assumed to have a multivariate normal distribution with variance-covariance matrix

$$\begin{pmatrix} 1 & \mathbf{B}_{uv}^d \\ \mathbf{B}_{uv}^{d\top} & \mathbf{B}_v^d \end{pmatrix}$$

where \mathbf{B}_{uv}^d and \mathbf{B}_v^d are unstructured. The variances of the u_i^d s are constrained to 1 so that the D processes have the same magnitude at baseline. Random effects are assumed independent between dimensions so that the entire $(D + q)$ -vector of individual random effects $\mathbf{w}_i = (u_i^\top, v_i^\top)^\top$ has a multivariate normal distribution,

$$\mathbf{w}_i \sim \mathcal{N}\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{I}_D & \mathbf{B}_{uv} \\ \mathbf{B}_{uv}^\top & \mathbf{B}_v \end{pmatrix}\right).$$

with \mathbf{I}_D the $D \times D$ identity matrix, \mathbf{B}_v the D -block diagonal matrix with d^{th} block \mathbf{B}_v^d , and \mathbf{B}_{uv} the $D \times q$ matrix with d^{th} row $\left(\mathbf{O}_{\sum_{l=1}^{d-1} q_l}, \mathbf{B}_{uv}^d, \mathbf{O}_{\sum_{l=d+1}^D q_l}\right)$ where \mathbf{O}_x is the x -row vector of zeros. In the estimation process, \mathbf{B} is replaced by its Cholesky decomposition: $\mathbf{B} = \mathbf{L}\mathbf{L}^\top$, where \mathbf{L} is a $(D + q) \times (D + q)$ lower triangular matrix. Independence between random effects across dimensions is handled by fixing corresponding parameters to zero in \mathbf{L} , and unit-variances of random intercepts are handled by fixing the corresponding parameters to 1 in \mathbf{L} . In addition to the variances of random intercepts fixed at one,

and without loss of generality, the matrix of covariates \mathbf{X}_i^0 excludes intercepts for each process so that the D processes are standardized at baseline. These two constraints are compensated by parameters in the measurement models.

The temporal influences between processes are modelled through the $D \times D$ -matrix of time-dependent transition intensities $\mathbf{A}_{i,\delta}(t)$:

$$\mathbf{A}_{i,\delta}(t) = \begin{pmatrix} a_{i,11}(t) & \dots & a_{i,1d}(t) & \dots & a_{i,1D}(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,d1}(t) & \dots & a_{i,dd}(t) & \dots & a_{i,dD}(t) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{i,D1}(t) & \dots & a_{i,Dd}(t) & \dots & a_{i,DD}(t) \end{pmatrix}$$

This matrix captures the directed temporal influences between latent processes at time t and subsequent rates of change of latent processes between times t and $t + \delta$. Specifically, coefficient $a_{i,dd'}(t)$ quantifies the temporal effect of process d' at time t on process d . Each effect can be modelled according to time/covariates through a linear regression $a_{i,dd'}(t) = \mathbf{R}_i^\top(t) \boldsymbol{\alpha}_{dd'}$ where $\mathbf{R}_i(t)$ is a r -vector of time-dependent covariates associated with the r -vector of regression coefficients $\boldsymbol{\alpha}_{dd'}^\top = (\alpha_{dd'}^m)_{m=0,(r-1)}^\top$.

When the discretization step is not too large, the temporal influences intend to have the same causal interpretations as those of a model in continuous time (see Simulation Study 2 for the demonstration).

3.2 Measurement Models of the longitudinal markers

Consider K ($K \geq D$) continuous longitudinal markers $\mathbf{Y}_{ij} = (Y_{ijk})_{k=1,\dots,K}^\top$ that have been measured for subject i at $(n_i + 1)$ successive discrete times $t_{ij} = j \times \delta$ with $j \in \tau_i = \{j_{0i}, \dots, (j_{0i} + n_i)\}$. Following Proust-Lima et al. (2006, 2013), we assume that each latent process Λ_i^d is the underlying common factor of K_d markers ($K = \sum_{d=1}^D K_d$) and we note \mathcal{K}^d the set of marker subscripts associated with latent process d . We assume that a marker measures only one latent process.

The link between a marker and its underlying latent process is defined by a marker-specific measurement model. If marker k is Gaussian, the measurement model is a linear equation:

$$\frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} = \tilde{Y}_{ijk} = \Lambda_i^d(t_{ij}) + \tilde{\epsilon}_{ijk}, \quad \forall k \in \mathcal{K}^d \ \& \ \forall j \in \tau_i \quad (2)$$

where the vector of transformation parameters $\boldsymbol{\eta}_k = (\eta_{0k}, \eta_{1k})^\top$ is used to get the standardized form \tilde{Y}_{ijk} of the marker and $\tilde{\epsilon}_{ijk}$ are independent Gaussian errors with variance σ_k^2 .

In the more general case of a continuous marker (possibly non Gaussian), one may consider a nonlinear equation of observation:

$$H_k(Y_{ijk}; \boldsymbol{\eta}_k) = \tilde{Y}_{ijk} = \Lambda_i^d(t_{ij}) + \tilde{\epsilon}_{ijk}, \quad \forall k \in \mathcal{K}^d \ \& \ \forall j \in \tau_i \quad (3)$$

where the link transformation H_k comes from a family of monotonic increasing and continuous functions parameterized with $\boldsymbol{\eta}_k$. Again \tilde{Y}_{ijk} is the transformed marker and $\tilde{\epsilon}_{ijk}$ are independent Gaussian errors with variance σ_k^2 . Following Proust-Lima et al. (2013), the link transformation H_k can be defined from a basis of I-splines (which are integrated M-splines (Ramsay, 1988)) in association with positive coefficients, thus providing an increasing bijective flexible transformation. We used here a quadratic I-splines basis with p_k internal knots, $(\mathbb{I}_m)_{m=1, p_k+3}$, so that

$$H_k(Y_{ijk}; \boldsymbol{\eta}_k) = \tilde{Y}_{ijk} = \eta_{0k} + \sum_{m=1}^{p_k+3} \eta_{mk}^2 \mathbb{I}_m(Y_{ijk}),$$

with $(\eta_{mk})_{m=0, p_k+3}$ the vector of parameters of the transformation.

In the following, we denote $\boldsymbol{\Sigma} = \text{diag}((\sigma_k^2)_{k=1, \dots, K})$ the diagonal variance matrix of the vector of errors $\tilde{\boldsymbol{\epsilon}}_{ij} = (\tilde{\epsilon}_{ijk})_{k=1, \dots, K}^\top$ and $\boldsymbol{\eta} = (\boldsymbol{\eta}_k^\top)_{\{1, \dots, K\}}^\top$, the total vector of transformation parameters for the K markers. The vector of transformed markers $\tilde{\mathbf{Y}}_{ij}$ is mapped to the system of latent processes $\boldsymbol{\Lambda}_i(t_{ij})$ through a $K \times D$ matrix \mathbf{P} with element (k, d) equal to 1 if marker k measures latent process d :

$$\tilde{\mathbf{Y}}_{ij} = \mathbf{P} \boldsymbol{\Lambda}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij} \quad (4)$$

In practice the observation process may include intermittent missing observations for a subset of markers or for all the markers at a given occasion $j \in \tau_i$, so that $K_{ij}^* \leq K$ markers are actually observed at occasion j for subject i . We assume that observations are missing at random and note $\tilde{\mathbf{Y}}_{ij}^*$ the transformations of the actual K_{ij}^* -vector of observed markers \mathbf{Y}_{ij}^* at occasion j .

The model linking the system of processes $\boldsymbol{\Lambda}_i(t_{ij})$ to the transformed markers $\tilde{\mathbf{Y}}_{ij}^*$ can be easily adapted to the presence of intermittent missing data by considering a $K_{ij}^* \times K$ observation matrix \mathbf{M}_{ij} (for $j \in \tau_i$) where element (k^*, k) equals 1 if marker k is the $k^{*\text{th}}$ observed marker at occasion j and 0 if not for $k = 1, \dots, K$ and $k^* = 1, \dots, K_{ij}^*$:

$$\tilde{\mathbf{Y}}_{ij}^* = \mathbf{M}_{ij} \mathbf{P} \boldsymbol{\Lambda}_i(t_{ij}) + \boldsymbol{\epsilon}_{ij}^* \quad (5)$$

with $\boldsymbol{\epsilon}_{ij}^*$ the vector of independent Gaussian errors with covariance matrix $\mathbf{M}_{ij}\boldsymbol{\Sigma}\mathbf{M}_{ij}^\top$.

3.3 Estimation

The parameters are estimated in the maximum likelihood framework.

3.3.1 Distribution of the Latent Processes and transformed observations

We first derive the marginal distributions of the latent processes and of the transformed observations. By recurrence, the structural model (1) can be rewritten:

$$\boldsymbol{\Lambda}_i(t_{ij}) = \begin{cases} \mathbf{X}_i^0\boldsymbol{\beta} + \mathbf{u}_i & \text{if } j = 0 \\ \prod_{l=0}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_{il})(\mathbf{X}_i^0\boldsymbol{\beta} + \mathbf{u}_i) & \text{if } j > 0 \\ +\delta \sum_{s=1}^j \prod_{l=s}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_{il})(\mathbf{X}_i(t_{is})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{is})\mathbf{v}_i) & \end{cases} \quad (6)$$

where $t_{ij} = j \times \delta$ for $j \in \{0, \dots, J\}$ and $\tilde{\mathbf{A}}_{i,\delta}(t_{ij}) = \mathbf{I}_D + \delta\mathbf{A}_{i,\delta}(t_{ij})$.

By introducing $\boldsymbol{\Psi}_{i,\delta}(t_0, j, s)$ for $t_0 \geq 0$ and $s \leq j$ so that

$$\boldsymbol{\Psi}_{i,\delta}(t_0, j, s) = \begin{cases} \mathbf{I}_D, & \text{if } s = j \\ \prod_{l=s}^{j-1} \tilde{\mathbf{A}}_{i,\delta}(t_0 + t_{il}) & \text{if } s < j \end{cases} \quad (7)$$

Equation (6) can be rewritten

$$\boldsymbol{\Lambda}_i(t_{ij}) = \boldsymbol{\Psi}_{i,\delta}(0, j, 0)(\mathbf{X}_i^0\boldsymbol{\beta} + \mathbf{u}_i) + \left(\delta \sum_{s=1}^j \boldsymbol{\Psi}_{i,\delta}(0, j, s)(\mathbf{X}_i(t_{is})\boldsymbol{\gamma} + \mathbf{Z}_i(t_{is})\mathbf{v}_i) \right) \mathbf{1}_{\{j>0\}} \quad (8)$$

As a result, the vector $\boldsymbol{\Lambda}_i(t_{ij})$ has a multivariate normal distribution with expectation $\boldsymbol{\mu}_{\boldsymbol{\Lambda}_{ij}}$ and variance covariance matrix $\mathbf{V}_{\boldsymbol{\Lambda}_{ij}} = \text{var}(\boldsymbol{\Lambda}_i(t_{ij}))$ and the vector $\boldsymbol{\Lambda}_i = (\boldsymbol{\Lambda}_i(t_{ij}))_{j \in \tau_i}^\top$ has a multivariate normal distribution with expectation $\boldsymbol{\mu}_{\boldsymbol{\Lambda}_i} = \left(\boldsymbol{\mu}_{\boldsymbol{\Lambda}_{ij}}^\top \right)_{j \in \tau_i}^\top$ and variance-covariance matrix $\mathbf{V}_{\boldsymbol{\Lambda}_i} = \left(\mathbf{V}_{\boldsymbol{\Lambda}_{ijj'}} \right)_{(j,j') \in \tau_i^2}$ where

$$\boldsymbol{\mu}_{\boldsymbol{\Lambda}_{ij}} = E[\boldsymbol{\Lambda}_i(t_{ij})] = \boldsymbol{\Psi}_{i,\delta}(0, j, 0)\mathbf{X}_i^0\boldsymbol{\beta} + \left(\delta \sum_{s=1}^j \boldsymbol{\Psi}_{i,\delta}(0, j, s)\mathbf{X}_i(t_{is})\boldsymbol{\gamma} \right) \mathbf{1}_{\{j>0\}} \quad (9)$$

and

$$\begin{aligned}
\mathbf{V}_{\Lambda_{ijj'}} &= \text{cov}(\Lambda_i(t_{ij}); \Lambda_i(t_{ij'})) \\
&= \Psi_{i,\delta}(0, j, 0) \Psi_{i,\delta}(0, j', 0)^\top \\
&\quad + \left(\Psi_{i,\delta}(0, j, 0) \mathbf{B}_{uv} \left(\delta \sum_{s'=1}^{j'} \Psi_{i,\delta}(0, j', s') \mathbf{Z}_i(t_{is'}) \right)^\top \right) \mathbb{1}_{\{j'>0\}} \\
&\quad + \left(\left(\delta \sum_{s=1}^j \Psi_{i,\delta}(0, j, s) \mathbf{Z}_i(t_{is}) \right) \mathbf{B}_{uv}^\top \Psi_{i,\delta}(0, j', 0)^\top \right) \mathbb{1}_{\{j>0\}} \\
&\quad + \left(\left(\delta \sum_{s=1}^j \Psi_{i,\delta}(0, j, s) \mathbf{Z}_i(t_{is}) \right) \mathbf{B}_v \left(\delta \sum_{s'=1}^{j'} \Psi_{i,\delta}(0, j', s') \mathbf{Z}_i(t_{is'}) \right)^\top \right) \mathbb{1}_{\{\min(j,j')>0\}}
\end{aligned} \tag{10}$$

It can be easily deduced that the vector of incomplete and transformed data $\tilde{\mathbf{Y}}_{ij}^*$ at occasion j is multivariate Gaussian with expectation $\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_{ij}^*} = \mathbf{M}_{ij} \mathbf{P} \boldsymbol{\mu}_{\Lambda_{ij}}$ and variance-covariance matrix $\mathbf{V}_{\tilde{\mathbf{Y}}_{ij}^*} = \mathbf{M}_{ij} (\mathbf{P} \mathbf{V}_{\Lambda_{ijj}} \mathbf{P}^\top + \boldsymbol{\Sigma}) \mathbf{M}_{ij}^\top$, and the total vector of incomplete and transformed data $\tilde{\mathbf{Y}}_i^* = \left(\tilde{\mathbf{Y}}_{ij}^{*\top} \right)_{j \in \tau_i}^\top$ is multivariate Gaussian with expectation $\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^*} = \left(\boldsymbol{\mu}_{\tilde{\mathbf{Y}}_{ij}^*}^\top \right)_{j \in \tau_i}^\top$ and variance-covariance matrix $\mathbf{V}_{\tilde{\mathbf{Y}}_i^*}$, a block matrix with $\mathbf{M}_{ij} \mathbf{P} \mathbf{V}_{\Lambda_{ijj'}} \mathbf{P}^\top \mathbf{M}_{ij'}^\top + \left(\mathbf{M}_{ij} \boldsymbol{\Sigma} \mathbf{M}_{ij'}^\top \right) \mathbb{1}_{\{j=j'\}}$ the (j, j') block.

3.3.2 Likelihood

As the N subjects of the sample are independent, the log-likelihood of the model is $L(Y^*; \boldsymbol{\theta}) = \sum_{i=1}^N \log(\mathcal{L}_i(\mathbf{Y}_i^*; \boldsymbol{\theta}))$ with $\mathcal{L}_i(\mathbf{Y}_i^*; \boldsymbol{\theta})$ the individual contribution to the likelihood. Here, $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \text{vec}(\mathbf{B})^\top, (\boldsymbol{\alpha}_{dd'}^\top)_{d,d' \in \{1, \dots, D\}^2}, (\sigma_k)_{k \in \{1, \dots, K\}}, \boldsymbol{\eta}^\top)^\top$ is the whole vector of parameters. Using the Jacobian of the link functions H_k ($k = 1, \dots, K$), the individual contribution is:

$$\mathcal{L}_i(\mathbf{Y}_i^*; \boldsymbol{\theta}) = \phi_i(\tilde{\mathbf{Y}}_i^*; \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^*}, \mathbf{V}_{\tilde{\mathbf{Y}}_i^*}) \prod_{j \in \tau_i} \prod_{l=1}^{K_{ij}^*} \mathcal{J}_{\kappa(l)} \left(H_{\kappa(l)}(\tilde{\mathbf{Y}}_{ij\kappa(l)}^*; \boldsymbol{\eta}_{\kappa(l)}) \right) \tag{11}$$

where $\phi_i(\cdot; \boldsymbol{\mu}, \mathbf{V})$ denotes the density function of a multivariate Gaussian vector with expectation $\boldsymbol{\mu}$ and variance-covariance matrix \mathbf{V} , and $\mathcal{J}_{\kappa(l)}(H_{\kappa(l)}(\tilde{\mathbf{Y}}_{ij\kappa(l)}^*; \boldsymbol{\eta}_{\kappa(l)}))$ denotes the Jacobian of the link function $H_{\kappa(l)}$ used to transform $\tilde{\mathbf{Y}}_{ij\kappa(l)}^*$, the l^{th} observed marker at oc-

caison j for subject i . For instance, with a linear link function, $\mathcal{J}_{\kappa(l)} \left(H_{\kappa(l)}(\tilde{Y}_{ij\kappa(l)}^*; \eta_{\kappa(l)}) \right) = \frac{1}{\eta_{1\kappa(l)}}$.

3.3.3 Optimisation Algorithm and Implementation

The maximum likelihood estimates are obtained using an extended Levenberg-Marquardt algorithm (Marquardt, 1963) because of its robustness and good convergence rate. At each iteration p , if necessary, the Hessian matrix $\mathbf{H}^{(p)}$ is diagonal-inflated to obtain a positive definite matrix $\mathbf{H}^{*(p)}$ which is used to update the parameters $\boldsymbol{\theta}^{(p+1)} = \boldsymbol{\theta}^{(p)} - \nu[\mathbf{H}^{*(p)}]^{-1}\mathbf{U}(\boldsymbol{\theta}^{(p)})$, with $\mathbf{U}(\boldsymbol{\theta}^{(p)})$ the gradient at iteration p and ν the improvement control parameter. Convergence is reached when $\|\boldsymbol{\theta}^{(p+1)} - \boldsymbol{\theta}^{(p)}\|_2 < 10^{-3}$, $|L(\mathbf{Y}^*; \boldsymbol{\theta}^{(p+1)}) - L(\mathbf{Y}^*; \boldsymbol{\theta}^{(p)})| < 10^{-3}$ and $\frac{\mathbf{U}(\boldsymbol{\theta}^{(p)})^\top [\mathbf{H}^{(p)}]^{-1} \mathbf{U}(\boldsymbol{\theta}^{(p)})}{n_{\text{para}}} < 10^{-2}$, with n_{para} the total number of parameters. The latter criterion is by far the most stringent one and specifically targets maximum search. The variances of the estimators are obtained from the inverse of $\mathbf{H}^{(p)}$.

Given the possibly high number of parameters, we first estimate the parameters for each process taken separately, then we start the maximization of the likelihood of the multivariate model from these simple estimates, setting initial values of the inter-dimension parameters to zero. The model estimation is implemented in R (program available on request); it combines R and C++ languages, and includes parallel computations.

3.4 Marginal and subject-specific predictions

The goodness-of-fit of the model can be assessed by comparing predictions with observations of the markers in their transformed scales. From notations defined in section 3.3.1, marginal and conditional distributions of the markers are:

$$\tilde{\mathbf{Y}}_i \sim \mathcal{N}(\mathbf{P}\boldsymbol{\mu}_{\Lambda_i}, (\mathbf{P}\mathbf{V}_{\Lambda_i}\mathbf{P}^\top + \boldsymbol{\Sigma}_i)). \quad (12)$$

$$\tilde{\mathbf{Y}}_i \Big|_{\Lambda_i} \sim \mathcal{N}(\mathbf{P}\boldsymbol{\Lambda}_i, \boldsymbol{\Sigma}_i) \quad (13)$$

where $\boldsymbol{\Sigma}_i$ is the block-diagonal matrix constituted of n_i $\boldsymbol{\Sigma}$ blocks.

The marginal ($\mathbf{Y}_i^{(M)}$) and subject-specific ($\mathbf{Y}_i^{(SS)}$) predictions in the transformed scales are then respectively obtained by taking the expectations of the marginal and conditional distributions of the transformed markers at the parameter estimates $\hat{\boldsymbol{\theta}}$ and at the predicted latent processes $\hat{\boldsymbol{\Lambda}}_i$ of $\boldsymbol{\Lambda}_i$ given the observations $\tilde{\mathbf{Y}}_i^*$: $\hat{\boldsymbol{\Lambda}}_i = E \left[\boldsymbol{\Lambda}_i | \tilde{\mathbf{Y}}_i^* \right] =$

$\boldsymbol{\mu}_{\Lambda_i} + \mathbf{C}_{\Lambda_i \tilde{\mathbf{Y}}_i^*} \mathbf{V}_{\tilde{\mathbf{Y}}_i^*}^{-1} (\tilde{\mathbf{Y}}_i^* - \boldsymbol{\mu}_{\tilde{\mathbf{Y}}_i^*})$, where $\mathbf{C}_{\Lambda_i \tilde{\mathbf{Y}}_i^*} = \left(\mathbf{V}_{\Lambda_{ijj'}} \mathbf{P}^\top \mathbf{M}_{ij'}^\top \right)_{(j,j') \in \tau_i^2}$ is the covariance matrix between Λ_i and $\tilde{\mathbf{Y}}_i^*$.

Using these individual predictions, one can graphically compare either the marginal predictions $\mathbf{Y}^{(M)}$ or subject-specific predictions $\mathbf{Y}^{(SS)}$ averaged within time intervals to the observations averaged within the same time intervals. Marginal and subject-specific predictions in the natural scale of the markers can also be derived from the marginal and conditional distributions by using a Monte-Carlo approximation (Proust-Lima et al., 2013).

4 Simulations

We performed two series of simulations to evaluate the estimation program and the impact of time discretization on the interpretation of $\mathbf{A}(t)$ matrix.

4.1 Simulation Study 1: Validation of the estimation program

4.1.1 Design

To evaluate the estimation program, we generated a system of two Gaussian processes ($(\Lambda^1(t))_{t \geq 0}$ and $(\Lambda^2(t))_{t \geq 0}$), each one measured by one longitudinal marker (\mathbf{Y}_1 and \mathbf{Y}_2) and two covariates, one binary C_1 and one continuous C_2 . We considered two scenarios: a covariate-specific temporal influence structure (Scenario 1) and a time-dependent temporal influence structure (Scenario 2). In Scenario 1, we assumed linear trajectories over time for the system of latent processes (random intercepts and simple effects of both covariates in the sub-models for the initial level and the change over time) and a temporal influence structure $\mathbf{A}(t)$ different in each level of C_1 :

$$\begin{cases} \Lambda_i^1(0) &= \beta_0^1 + \beta_1^1 C_{1,i} + \beta_2^1 C_{2,i} + u_i^1 \\ \Lambda_i^2(0) &= \beta_0^2 + \beta_1^2 C_{1,i} + \beta_2^2 C_{2,i} + u_i^2 \\ \frac{\Delta \Lambda_i^1(t+\delta)}{\delta} &= \gamma_0^1 + \gamma_1^1 C_{1,i} + \gamma_2^1 C_{2,i} + v_i^1 + (\alpha_{i,11}^0 + \alpha_{i,11}^1 C_{1,i}) \Lambda_i^1(t) + (\alpha_{i,12}^0 + \alpha_{i,12}^1 C_{1,i}) \Lambda_i^2(t) \\ \frac{\Delta \Lambda_i^2(t+\delta)}{\delta} &= \gamma_0^2 + \gamma_1^2 C_{1,i} + \gamma_2^2 C_{2,i} + v_i^2 + (\alpha_{i,21}^0 + \alpha_{i,21}^1 C_{1,i}) \Lambda_i^1(t) + (\alpha_{i,22}^0 + \alpha_{i,22}^1 C_{1,i}) \Lambda_i^2(t) \\ \frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} &= \Lambda_{ij}^k + \tilde{\epsilon}_{ijk}, \quad k = 1, 2 \end{cases} \quad (14)$$

where $\mathbf{u}_i = (u_i^1, u_i^2)^\top$, $\mathbf{v}_i = (v_i^1, v_i^2)^\top$ and $(u_i^\top, v_i^\top)^\top \sim \mathcal{N}(0, \mathbf{L}\mathbf{L}^\top)$ with \mathbf{L} such that the random effects are independent between dimensions, and $\tilde{\epsilon}_{ijk} \sim \mathcal{N}(0, \sigma_k^2)$, $\forall k \in \{1, 2\}$.

In scenario 2, we assumed linear trajectories for the latent processes adjusted for C_2 and a transition matrix that evolved with time:

$$\left\{ \begin{array}{l} \Lambda_i^1(0) = \beta_0^1 + \beta_1^1 C_{2,i} + u_i^1 \\ \Lambda_i^2(0) = \beta_0^2 + \beta_1^2 C_{2,i} + u_i^2 \\ \frac{\Delta \Lambda_i^1(t+\delta)}{\delta} = \gamma_0^1 + v_i^1 + a_{11}(t) \Lambda_i^1(t) + a_{12}(t) \Lambda_i^2(t) \\ \frac{\Delta \Lambda_i^2(t+\delta)}{\delta} = \gamma_0^2 + v_i^2 + a_{21}(t) \Lambda_i^1(t) + a_{22}(t) \Lambda_i^2(t) \\ \frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} = \Lambda_{ij}^k + \tilde{\epsilon}_{ijk}, \quad k = 1, 2, \end{array} \right. \quad (15)$$

where $\mathbf{u}_i = (u_i^1, u_i^2)^\top$, $\mathbf{v}_i = (v_i^1, v_i^2)^\top$ and $(u_i^\top, v_i^\top)^\top \sim \mathcal{N}(0, \mathbf{L}\mathbf{L}^\top)$ with \mathbf{L} such that the random effects are independent between dimensions, and $\tilde{\epsilon}_{ijk} \sim \mathcal{N}(0, \sigma_k^2)$, $\forall k \in \{1, 2\}$. Each element of the transition matrix $a_{kk'}(t)$ is defined from a basis of quadratic B-splines with one internal knot at the median $(\mathbf{S}_m)_{\{m=1,3\}}$ so that $a_{kk'}(t) = \alpha_{kk'}^0 + \alpha_{kk'}^1 \mathbf{S}_1(t) + \alpha_{kk'}^2 \mathbf{S}_2(t) + \alpha_{kk'}^3 \mathbf{S}_3(t)$, $\forall k, k' \in \{1, 2\}^2$.

The design of the simulations and the parameters were chosen to mimic the ADNI data. Dimensions 1 and 2 were cerebral anatomy and cognitive ability, respectively measured by a specific composite score. Covariate C_1 corresponded to the two groups CN (for normal elderly, in reference) and MCI (for Mild Cognitive Impairment). It was generated according to a Bernoulli distribution with probability $P(C_1 = 1) = 0.64$. Covariate C_2 corresponded to the standardized age at baseline and was generated according to a standard Gaussian distribution. We considered a discretization step δ of 6 months. Markers observations were generated every 6 months up to 3 years. Thus a subject had 7 repeated measures at occasions $j \in \{0, 1, \dots, 6\}$. We also considered a design in which scheduled visits could be missed completely at random with a probability of 0.15, and when a visit was not missed, a marker could be missing with a probability of 0.07. For each design and scenario, we generated 100 samples of 512 subjects.

4.1.2 Results

Table 4 and Table 1 provide the results of the simulations for scenarios 1 and 2 respectively. In both settings, all the parameters were correctly estimated with satisfying coverage rates in the absence of missing data (left part) and in the presence of missing data (right part).

Table 1: Results of the simulations for scenario 2 (100 replicates of samples of size 512).

| | θ | without missing values (conv [†] =100%) | | | | with missing values * (conv [†] =98%) | | | |
|-----------------|----------|--|------------------|------------------|-------|--|------------------|------------------|-------|
| | | $\hat{\theta}$ | ESE [‡] | ASE [‡] | CR(%) | $\hat{\theta}$ | ESE [‡] | ASE [‡] | CR(%) |
| β_1^1 | -1.661 | -1.655 | 0.104 | 0.112 | 96.0 | -1.653 | 0.111 | 0.113 | 98.0 |
| β_2^1 | -1.803 | -1.831 | 0.146 | 0.128 | 92.0 | -1.834 | 0.150 | 0.133 | 90.0 |
| γ_0^1 | 0.278 | 0.281 | 0.043 | 0.046 | 96.0 | 0.285 | 0.045 | 0.050 | 98.0 |
| γ_1^1 | -0.152 | -0.154 | 0.023 | 0.023 | 96.0 | -0.156 | 0.024 | 0.026 | 96.0 |
| γ_0^2 | 0.117 | 0.121 | 0.080 | 0.080 | 94.0 | 0.111 | 0.093 | 0.086 | 91.0 |
| γ_1^2 | -0.143 | -0.145 | 0.040 | 0.041 | 96.0 | -0.140 | 0.046 | 0.045 | 95.0 |
| L(3,1) | 0.041 | 0.040 | 0.019 | 0.022 | 97.0 | 0.039 | 0.020 | 0.024 | 98.9 |
| L(4,2) | 0.012 | 0.026 | 0.021 | 0.039 | 97.0 | 0.029 | 0.025 | 0.044 | 97.0 |
| L(3,3) | 0.198 | 0.197 | 0.015 | 0.014 | 94.0 | 0.196 | 0.016 | 0.014 | 92.0 |
| L(4,4) | 0.373 | 0.372 | 0.027 | 0.027 | 95.0 | 0.371 | 0.029 | 0.029 | 93.0 |
| α_{11}^1 | 0.112 | 0.109 | 0.040 | 0.044 | 96.0 | 0.107 | 0.046 | 0.048 | 97.0 |
| α_{11}^2 | 0.003 | 0.010 | 0.047 | 0.050 | 95.0 | 0.015 | 0.052 | 0.056 | 96.0 |
| α_{11}^3 | -0.033 | -0.050 | 0.070 | 0.070 | 97.0 | -0.056 | 0.070 | 0.079 | 98.0 |
| α_{11}^4 | -0.132 | -0.117 | 0.095 | 0.093 | 95.0 | -0.108 | 0.099 | 0.104 | 97.0 |
| α_{12}^1 | 0.168 | 0.170 | 0.039 | 0.037 | 93.0 | 0.172 | 0.044 | 0.041 | 92.0 |
| α_{12}^2 | 0.079 | 0.067 | 0.043 | 0.043 | 92.0 | 0.067 | 0.049 | 0.048 | 92.0 |
| α_{12}^3 | 0.143 | 0.160 | 0.062 | 0.057 | 92.0 | 0.161 | 0.065 | 0.064 | 92.0 |
| α_{12}^4 | 0.070 | 0.050 | 0.080 | 0.073 | 90.0 | 0.048 | 0.088 | 0.081 | 89.0 |
| α_{21}^1 | 0.238 | 0.225 | 0.071 | 0.068 | 95.0 | 0.220 | 0.076 | 0.073 | 92.0 |
| α_{21}^2 | 0.138 | 0.159 | 0.092 | 0.090 | 94.0 | 0.162 | 0.097 | 0.097 | 93.0 |
| α_{21}^3 | 0.359 | 0.331 | 0.128 | 0.126 | 94.0 | 0.330 | 0.134 | 0.135 | 95.0 |
| α_{21}^4 | 0.014 | 0.048 | 0.170 | 0.165 | 94.0 | 0.045 | 0.179 | 0.178 | 95.0 |
| α_{22}^1 | 0.108 | 0.108 | 0.077 | 0.074 | 91.0 | 0.102 | 0.085 | 0.081 | 89.0 |
| α_{22}^2 | 0.037 | 0.024 | 0.082 | 0.084 | 97.0 | 0.025 | 0.091 | 0.088 | 95.0 |
| α_{22}^3 | -0.067 | -0.046 | 0.105 | 0.107 | 90.0 | -0.049 | 0.116 | 0.113 | 91.0 |
| α_{22}^4 | -0.066 | -0.093 | 0.133 | 0.133 | 96.0 | -0.085 | 0.154 | 0.142 | 94.0 |
| η_{01} | 0.387 | 0.391 | 0.015 | 0.015 | 93.0 | 0.391 | 0.015 | 0.016 | 96.0 |
| η_{02} | 0.713 | 0.713 | 0.035 | 0.033 | 94.0 | 0.709 | 0.035 | 0.035 | 95.0 |
| η_{11} | 3.769 | 3.725 | 0.192 | 0.197 | 95.0 | 3.719 | 0.199 | 0.199 | 93.0 |
| η_{12} | 2.580 | 2.561 | 0.092 | 0.090 | 95.0 | 2.563 | 0.094 | 0.092 | 95.0 |
| σ_1 | 2.489 | 2.511 | 0.138 | 0.120 | 92.0 | 2.515 | 0.144 | 0.124 | 91.0 |
| σ_2 | 1.412 | 1.413 | 0.063 | 0.060 | 93.0 | 1.415 | 0.063 | 0.064 | 95.0 |

[†] Rate of convergence,

* (15% missing occasions, 7% missing outcomes),

[‡] ASE is the asymptotic standard error, ESE is the empirical standard error and CR is the coverage rate of the 95% confidence interval.

4.2 Simulation Study 2: Impact of the discretization step on the temporal influence structure between processes

4.2.1 Design

To formally assess whether the interpretation under the discretized time was the same as the one obtained under continuous time, we assessed the type-I error rate associated with each non diagonal element of the transition matrix \mathbf{A} under three discretization steps $\delta = 1/3, 1/2$ and 1 when data were actually generated under continuous time (approximated by a step $\delta=0.001$). We considered for this a system of three latent processes ($(\mathbf{\Lambda}^1(t))_{t \geq 0}$, $(\mathbf{\Lambda}^2(t))_{t \geq 0}$, and $(\mathbf{\Lambda}^3(t))_{t \geq 0}$), each one measured by one Gaussian repeated marker ($\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3$). The trajectories were linear, with no adjustment for covariates and a transition matrix \mathbf{A} constant over time:

$$\left\{ \begin{array}{l} \Lambda_i^1(0) = \beta_0^1 + u_i^1 \\ \Lambda_i^2(0) = \beta_0^2 + u_i^2 \\ \Lambda_i^3(0) = \beta_0^3 + u_i^3 \\ \frac{\Delta \Lambda_i^1(t+\delta)}{\delta} = \gamma_0^1 + v_i^1 + a_{11} \Lambda_i^1(t) + a_{12} \Lambda_i^2(t) + a_{13} \Lambda_i^3(t) \\ \frac{\Delta \Lambda_i^2(t+\delta)}{\delta} = \gamma_0^2 + v_i^2 + a_{21} \Lambda_i^1(t) + a_{22} \Lambda_i^2(t) + a_{23} \Lambda_i^3(t) \\ \frac{\Delta \Lambda_i^3(t+\delta)}{\delta} = \gamma_0^3 + v_i^3 + a_{31} \Lambda_i^1(t) + a_{32} \Lambda_i^2(t) + a_{33} \Lambda_i^3(t) \end{array} \right. \quad (16)$$

$$\frac{Y_{ijk} - \eta_{0k}}{\eta_{1k}} = \Lambda_{ij}^k + \tilde{\epsilon}_{ijk}, \quad k = 1, 2, 3$$

The simulation model mimicked the ADNI data with cerebral anatomy, cognitive ability and functional autonomy as dimensions, respectively measured by a specific composite score. We estimated the model on the ADNI data with $\delta=1$ (6 months) and transformed the parameters to the scale $\delta=0.001$ to generate the data in almost continuous time. We provide in the Appendix A: the formulas that we used to relate model components defined under $\delta=1$ and $\delta=0.001$. Elements of the transition matrix were set one by one to 0 in $\delta = 0.001$ scale to evaluate the associated type-I error rate. Latent processes were generated with solver from dsolve package (Soetaert et al., 2010) and observations were derived every 6 months up to 3 years. For each design, we considered 200 samples of 512 subjects.

4.2.2 Results

Table 2 displays the type-I error rates (in percentage) associated with each non diagonal element of the transition matrix \mathbf{A} when estimated with discretization steps $\delta=1/3, \delta=1/2$

and $\delta=1$. All the type-I error rates are close to the nominal 5% rate (expected rates in the 95% interval [1.98 , 8.02] with 200 replicates). This shows that with such discretization steps, the causal interpretation expected in continuous time is not altered.

Table 2: type-I error rates (in %) associated with each non diagonal element of the transition matrix \mathbf{A} when the transition matrix is generated approximately in continuous time ($\delta=0.001$) and estimated with discretization steps: $\delta=1/3$, $\delta=1/2$, $\delta=1$ (expected 95% interval [1.98 , 8.02] for the nominal type-I error of 5%).

| Parameter | $\delta = 1/3$ | $\delta = 1/2$ | $\delta = 1$ |
|-----------|----------------|----------------|--------------|
| a_{12} | 5.5 | 4.5 | 6.5 |
| a_{13} | 3.5 | 5.5 | 5.0 |
| a_{21} | 7.0 | 7.5 | 8.0 |
| a_{23} | 5.0 | 4.5 | 5.0 |
| a_{31} | 4.0 | 3.5 | 5.5 |
| a_{32} | 7.0 | 7.5 | 5.5 |

5 Application

The application aimed to describe the decline over time of cerebral anatomy, cognitive ability and functional autonomy in three clinical stages of AD (normal aging (CN), Mild Cognitive Impairment (MCI) and diagnosed with Alzheimer’s Disease (dAD)) and to quantify the temporal influences between these dimensions by assessing especially whether the relationships differed according to the clinical stage.

5.1 The sample

We considered the nineteen available cognitive tests as outcomes of global cognitive ability (Crane et al., 2012), the volume of hippocampus relative to total intracranial volume and the cortical thickness of nine regions (Freesurfer version 4.4.0 for longitudinal data) as outcomes for cerebral anatomy following regions of interest identified by Dickerson et al. (2008), and the sumscore of the FAQ composed of 30 items (Pfeffer et al., 1982) to assess functional autonomy. For numerical simplicity, we considered a unique marker for each dimension by summarizing the cerebral and cognitive markers into two composite scores (Proust-Lima et al., 2017). Covariates were the age at entry (centered around 75.4 and

indicated in decades), gender, educational level (low level if ≤ 12 years *vs* high level if >12 years), the APOE genotype ($\epsilon 4$ carrier *vs* $\epsilon 4$ non-carrier) and the 3 clinical stages (CN, MCI, dAD). We selected in the sample all the subjects who had no missing data for the covariates and had at least one measure for each dimension during the follow-up. The main analysis included 656 subjects (82% of the initial sample). The sample consisted of 190 (29%) subjects at healthy stage (CN), 322 (49%) subjects at mild cognitive impairment stage (MCI), and 144 (22%) subjects diagnosed with Alzheimer’s disease (dAD); 43% were females, 83% had a high educational level and 49% carried APOE $\epsilon 4$ allele. The mean age at entry was 75.4 years old (sd=6.6). The mean number of visits was 5, 6 and 4 for CN, MCI and dAD subjects, respectively.

5.2 The dynamic multivariate model

The dynamic model applied to ADNI is summarized in Figure 1. The three latent processes (cerebral anatomy (Λ^A), cognitive ability (Λ^C), functional autonomy (Λ^F)) were repeatedly measured by the three composite scores \mathbf{Y}_A , \mathbf{Y}_C and \mathbf{Y}_F . We assumed that changes of the processes were linear in time; thus, the processes had quadratic trajectories over time, adjusted for gender, education, APOE genotype, clinical stage, and age at baseline. In order to account for the correlation between individual repeated measures, we included random intercepts on the initial level of the processes and random intercepts and slopes on the change of the processes over time. Note that random effects on initial levels and on changes are correlated within each dimension but not between dimensions. The transition matrix that captured the interrelations between processes was constant over time and adjusted for clinical stage. To correct the possible departure from normality of the composite scores, we transformed them using a link function approximated by quadratic I-splines with 2 internal knots chosen at the terciles for cerebral and cognitive composite scores and by quadratic I-splines with one internal knot chosen at the median for the functional composite score.

In the main analysis, we took a 6 months discretization step. The strategy of analysis consisted in finding first the best adjustment for each process taken separately (assuming independency between processes ie., $a_{dd'}_{\{d \neq d'\}} = 0$ in Figure 1) with a significance threshold for covariate effects at 25%. Then, the whole multivariate model including all elements of the transition matrix was estimated. In secondary analyses, we re-estimated the final model by considering discretization steps of 3 months and of 2 months in order to evaluate whether the interpretations varied with smaller steps.

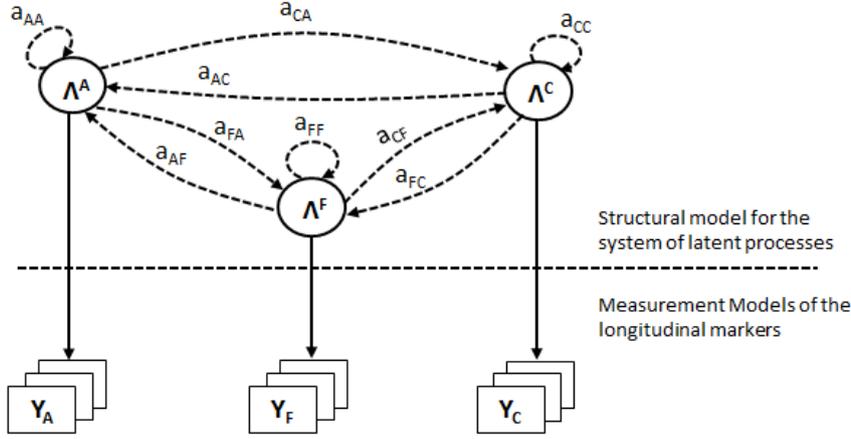


Figure 1: Graph of the dynamic causal model considered on ADNI data with three dimensions (labelled Λ^A for cerebral anatomy, Λ^C for cognitive ability and Λ^F for functional autonomy), each one repeatedly measured by one marker Y_A , Y_C and Y_F .

5.3 Results

5.3.1 Latent process-specific trajectories

Estimates of fixed effects, Cholesky's decomposition parameters (for the random effects covariance matrix) and measurement model parameters are provided in Tables 5, 6 and 7. In summary, older age, male gender, APOE $\epsilon 4$ carrying, and clinical stages MCI and dAD were associated with lower cerebral anatomy and lower cognitive ability levels at baseline. Higher education was associated with lower cerebral anatomy level but associated with higher cognitive ability level at baseline. Only clinical stages MCI and dAD were associated with lower functional autonomy level at baseline. APOE $\epsilon 4$ carrying was associated with steeper declines in cerebral anatomy and cognitive ability. Clinical stage dAD was associated with steeper cerebral anatomy decline and both clinical stages MCI and dAD were associated with steeper declines in cognitive ability and functional autonomy. Higher education was associated with smaller cognitive ability decline. Figure 4 depicts the expected trajectories of each dimension according to stage for two profiles of individuals (women noncarrier of APOE $\epsilon 4$ and with lower educational level; men carrier of APOE $\epsilon 4$ and with higher educational level).

5.3.2 Temporal influence structure between processes

Estimates of the transition intensity matrix are given in Table 3 and summarized in Figure 2. Figure 2 shows that the temporal influences between cerebral anatomy (Λ^A), cognitive ability (Λ^C), and functional autonomy (Λ^F) slightly evolve from healthy stage to Alzheimer's disease stage. At all stages, cerebral dimension affects significantly the dynamics of both cognitive and functional dimensions. Moreover, the cognitive and functional dimensions are interrelated, each one affecting the change over time of the other. However from MCI stage, the cognitive dimension becomes central by affecting significantly both the dynamics of cerebral and functional dimensions.

When considering a discretization step of 3 or 2 months instead of 6 months, the Akaike Information criterion was slightly improved (AIC=1386.68 with 6 months, AIC= 1370.45 with 3 months and AIC=1367.21 with 2 months) but the results regarding the temporal influence relationships between dimensions remained the same (see Table 8).

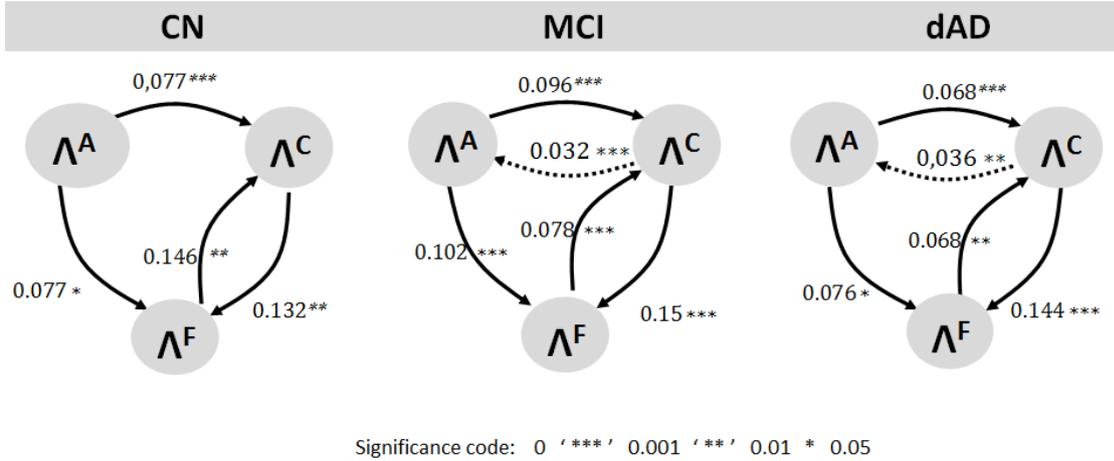


Figure 2: Temporal relationships estimated between cerebral anatomy (Λ^A), cognitive ability (Λ^C), and functional autonomy (Λ^F) at healthy (CN), MCI and dAD stages. Arrows represent effects of one dimension on the change of another dimension between two discretized times. Numbers indicate intensity effects, stars indicate the significance of the effects. Arrows in solid line indicate effects that exist at all stages; arrows in dashed line indicate effects that exist only for some stages.

Table 3: Estimates of the intensities of transition between cerebral anatomy, cognitive ability and functional autonomy using a discretization step of 6 months.

| | Parameter | | Estimate | SE | p-value |
|--|-----------|-----------------|----------|-------|---------|
| <u>Influence on Cerebral anatomy:</u> | | | | | |
| Effect of Cerebral dimension | Intercept | α_0^{11} | -0.054 | 0.014 | <0.001 |
| | MCI | α_1^{11} | 0.008 | 0.009 | 0.343 |
| | dAD | α_2^{11} | 0.001 | 0.013 | 0.971 |
| Effect of Cognitive dimension | Intercept | α_0^{12} | 0.010 | 0.008 | 0.196 |
| | MCI | α_1^{12} | 0.022 | 0.010 | 0.023 |
| | dAD | α_2^{12} | 0.025 | 0.014 | 0.076 |
| Effect of Functional dimension | Intercept | α_0^{13} | 0.010 | 0.017 | 0.557 |
| | MCI | α_1^{13} | -0.012 | 0.017 | 0.469 |
| | dAD | α_2^{13} | -0.011 | 0.019 | 0.576 |
| <u>Influence on Cognitive ability:</u> | | | | | |
| Effect of Cerebral dimension | Intercept | α_0^{21} | 0.077 | 0.023 | 0.001 |
| | MCI | α_1^{21} | 0.018 | 0.026 | 0.489 |
| | dAD | α_2^{21} | -0.009 | 0.034 | 0.784 |
| Effect of Cognitive dimension | Intercept | α_0^{22} | -0.396 | 0.074 | <0.001 |
| | MCI | α_1^{22} | 0.081 | 0.030 | 0.007 |
| | dAD | α_2^{22} | 0.026 | 0.042 | 0.526 |
| Effect of Functional dimension | Intercept | α_0^{23} | 0.146 | 0.054 | 0.007 |
| | MCI | α_1^{23} | -0.068 | 0.054 | 0.205 |
| | dAD | α_2^{23} | -0.078 | 0.059 | 0.184 |
| <u>Influence on Functional Autonomy:</u> | | | | | |
| Effect of Cerebral dimension | Intercept | α_0^{31} | 0.077 | 0.035 | 0.028 |
| | MCI | α_1^{31} | 0.025 | 0.043 | 0.558 |
| | dAD | α_2^{31} | -0.001 | 0.052 | 0.982 |
| Effect of Cognitive dimension | Intercept | α_0^{32} | 0.132 | 0.042 | 0.002 |
| | MCI | α_1^{32} | 0.018 | 0.048 | 0.707 |
| | dAD | α_2^{32} | 0.012 | 0.061 | 0.847 |
| Effect of Functional dimension | Intercept | α_0^{33} | -0.892 | 0.108 | <0.001 |
| | MCI | α_1^{33} | 0.569 | 0.078 | <0.001 |
| | dAD | α_2^{33} | 0.478 | 0.082 | <0.001 |

5.3.3 Goodness-of-fit of the model

We assessed the goodness-of-fit of the model by comparing the subject-specific predicted trajectories (conditionally to the random effects) with the observed trajectories in the transformed scale of the markers. Specifically, we computed the mean predictions and

mean transformed observations at each visit (every 6 months). As shown in Figure 3, the mean predictions are very close to the mean observations showing the good fit of the model to the data. These trajectories also illustrate the differences in trajectories according to clinical stage with each dimension more and more impaired and decline steeper and steeper with disease progression.

6 Discussion

We proposed an original dynamic model which aims to simultaneously describe multivariate processes over time and retrieve temporal relationships between the processes involved. Our model aims to be a dynamic causal model in the spirit of mechanistic models. We relied on discrete time with difference equations rather than continuous time with differential equations to largely reduce the numerical problems, notably with a closed-form likelihood.

We fundamentally explored in this work temporal influences between dynamic processes and the causal interpretation of these influences has to be done with caution. We relied on the dynamic approach to causality notably developed by Aalen and Frigessi (2007), Arjas and Parner (2004), Didelez (2008), Commenges and Gégout-Petit (2009). There is also a large literature in statistics on the potential outcomes formalism introduced by Rubin (2005); it is however not adapted for modeling dynamic relationships between non-directly observed factors, which is the aim in this paper so we do not discuss this approach further.

Following the dynamic approach to causality, we chose to analyze the changes of processes (as seeking local dependence structures) and defined causal relationships in a structural model, that is at a latent process level rather than at an observation level. One major simplifying assumption in our approach was the time discretization. Fundamentally, causal relationships are to be explored at an infinitesimal level and thus, a causal model is to be defined in continuous time (Commenges and Gégout-Petit, 2009; Aalen et al., 2016). To assess whether we could claim the same type of interpretation as mechanistic models, we thus assessed in a simulation study the impact of the discretization of the underlying time-continuous causal structure; we found that the type-I error rates of the temporal influence parameters were not altered by the use of a model in discrete time, provided the discretization step remained small in regards with the dynamics of the disease under study. Indeed in our simulations and application on Alzheimer’s disease, our

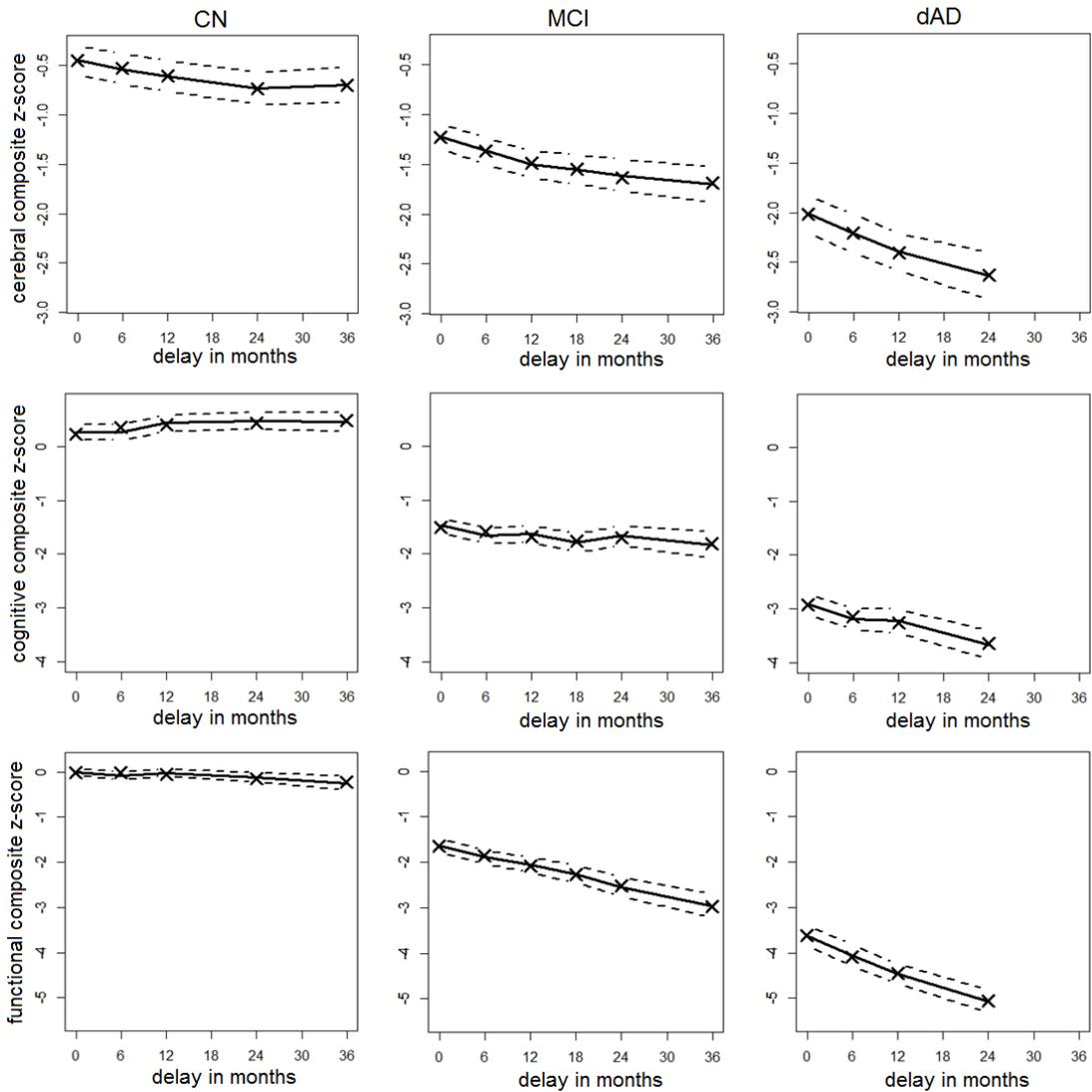


Figure 3: Means of subject-specific predictions at each visit (with crosses) along with the corresponding means of transformed observed scores (plain lines) and their 95% confidence interval (dashed lines). Columns refer to the group (controls CN, Subjects with Mild Cognitive Impairments MCI and subjects diagnosed with Alzheimer's disease dAD). Rows refer to the dimensions (cerebral anatomy, cognitive ability and functional autonomy).

discretization step was between 2 and 6 months while the disease progresses over decades (Amieva et al., 2008; Jack et al., 2013). In addition to the assessment that the absence of causal associations could be retrieved in the discretized time, we also provided in supplementary material analytic approximate relationships between transition matrices defined in two discretization steps or in continuous time and discrete time. These relationships quantify the associations between temporal influence structures; they remain however approximate as they were found under the assumption that the model structure did not differ substantially between the two discretization steps. This assumption deserves more attention in future work.

We acknowledge that despite our intention to provide a dynamic causal model, causal interpretation still has to be made cautiously as in any causal model since it is always subject to a correct specification of the system and of the statistical model.

The estimation we proposed relies on the assumption that missing data, both intermittent and monotone, are missing at random. This assumption is realistic in ADNI 1 which had a very short follow-up (3 years max). However, in applications with a longer follow-up, and notably in population-based cohorts, such assumption might become restrictive and it will be probably necessary to account for the occurrence of informative clinical events, such as dementia and death in our context. This will be possible by extending our model to a joint analysis of times to event but probably at the price of additional numerical complexities. In addition, in studies on chronic diseases with a long follow-up, it is reasonable to assume that the temporal influence structure may evolve with time. Our model can handle this as shown in the simulations where the temporal influence structure was modelled according to time using regression splines. We did not consider however time-dependent temporal influence structures in the application due to the short follow up. In addition, our objective was to contrast the influence structures between clinical stages, and as such we rather included stage-specific intensities of transition.

The model was primarily motivated by the study of the multiple alterations involved in the dementia process in the elderly. Indeed, dementia is characterized by very long and progressive alterations in different cerebral and clinical dimensions. However, although there is a global agreement on the relevant dimensions in AD, their relationships are still poorly understood and a confusion persists between alterations due to normal aging and alterations due to pathological aging leading to dementia. Thanks to the ADNI data which included subjects at different clinical stages of dementia, we were able to exhibit the global structure of dependence between cerebral, cognitive and functional domains

in normal aging characterized as expected by the cerebral structure which explains a part of the cognitive and functional changes, and the functional and cognitive changes which have bidirectional relationships. We also found out that the main evolution in this structure due to the pathological process towards dementia (in subjects with mild cognitive impairment or diagnosed with AD) was on cognitive functioning with some effect of cognitive functioning on change in cerebral structure in addition to its effect on change in functional structure. Such backward influence of cognition on anatomic structure may appear counterintuitive from a biological point of view. Yet, some studies showed that interventions made on cognitive functioning (e.g., literacy at adulthood) could induce an improvement in cerebral structure (Carreiras et al., 2009; Boyke et al., 2008). This application which gives a first insight on the possibilities of this dynamic model, could be now refined by targeting specific brain regions (separating for instance cortical thicknesses from hippocampal volume) and specific cognitive functions (separating for instance episodic memory from executive functioning).

As a conclusion, we proposed here a new methodology that may help identify temporal structures in multivariate longitudinal data. Although motivated and applied in dementia context, this approach has some potential to address unsolved questions in many other chronic diseases where multiple processes are in play, and or time-dependent exposures are to be assessed.

7 Acknowledgements

This work benefited from the support of the project SMALA (ANR-15-CE37-0002) of the French National Research Agency (ANR). Computer time was provided by the computing facilities MCIA (Mésocentre de Calcul Intensif Aquitain) at the Université de Bordeaux and the Université de Pau et des Pays de l'Adour.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

Aalen, O., Røysland, K., Gran, J., Kouyos, R., and Lange, T. (2016). Can we believe the dags? a comment on the relationship between causal dags and mechanisms. *Statistical*

methods in medical research **25**, 2294–2314.

- Aalen, O. O. and Frigessi, A. (2007). What can statistics contribute to a causal understanding? *Scandinavian Journal of Statistics* **34**, 155–168.
- Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., and Dartigues, J. F. (2008). Prodromal Alzheimer’s disease: successive emergence of the clinical symptoms. *Annals of neurology* **64**, 492–498.
- Arjas, E. and Parner, J. (2004). Causal reasoning from longitudinal data. *Scandinavian Journal of Statistics* **31**, 171–187.
- Boyke, J., Driemeyer, J., Gaser, C., Büchel, C., and May, A. (2008). Training-induced brain structure changes in the elderly. *Journal of Neuroscience* **28**, 7031–7035.
- Carreiras, M., Seghier, M. L., Baquero, S., Estévez, A., Lozano, A., Devlin, J. T., and Price, C. J. (2009). An anatomical signature for literacy. *Nature* **461**, 983.
- Commenges, D. and Gégout-Petit, A. (2009). A General Dynamical Statistical Model with Causal Interpretation. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **71**, 719–736.
- Crane, P. K., Carle, A., Gibbons, L. E., Insel, P., Mackin, R. S., Gross, A., Jones, R. N., Mukherjee, S., Curtis, S. M., Harvey, D., et al. (2012). Development and assessment of a composite score for memory in the alzheimer’s disease neuroimaging initiative (ADNI). *Brain imaging and behavior* **6**, 502–516.
- Desmée, S., Mentré, F., Veyrat-Follet, C., Sébastien, B., and Guedj, J. (2017). Nonlinear joint models for individual dynamic prediction of risk of death using Hamiltonian Monte Carlo: application to metastatic prostate cancer. *BMC Medical Research Methodology* **17**, 105.
- Dickerson, B. C., Bakkour, A., Salat, D. H., Feczko, E., Pacheco, J., Greve, D. N., Grodstein, F., Wright, C. I., Blacker, D., Rosas, H. D., et al. (2008). The cortical signature of Alzheimer’s disease: regionally specific cortical thinning relates to symptom severity in very mild to mild ad dementia and is detectable in asymptomatic amyloid-positive individuals. *Cerebral cortex* **19**, 497–510.

- Didelez, V. (2008). Graphical models for marked point processes based on local independence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **70**, 245–264.
- Donohue, M. C., Jacqmin-Gadda, H., Le Goff, M., Thomas, R. G., Raman, R., Gamst, A. C., Beckett, L. A., and et al. (2014). Estimating long-term multivariate progression from short-term data. *Alzheimer’s & Dementia: The Journal of the Alzheimer’s Association* **10**, S400–410.
- Greenland, S. (2000). Causal Analysis in the Health Sciences. *Journal of the American Statistical Association* **95**, 286–289.
- Hall, C. B., Lipton, R. B., Sliwinski, M., and Stewart, W. F. (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer’s disease. *Statistics in medicine* **19**, 1555–1566.
- Han, S. D., Gruhl, J., Beckett, L., Dodge, H. H., Stricker, N. H., Farias, S., Mungas, D., Initiative, A. D. N., et al. (2012). Beta amyloid, tau, neuroimaging, and cognition: sequence modeling of biomarkers for alzheimer’s disease. *Brain imaging and behavior* **6**, 610–620.
- Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., et al. (2013). Tracking pathophysiological processes in Alzheimer’s disease: an updated hypothetical model of dynamic biomarkers. *The Lancet Neurology* **12**, 207–216.
- Landau, S. M., Harvey, D., Madison, C. M., Koeppe, R. A., Reiman, E. M., Foster, N. L., Weiner, M. W., and Jagust, W. J. (2011). Associations between cognitive, functional, and fdg-pet measures of decline in AD and MCI. *Neurobiology of Aging* **32**, 1207 – 1218.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics* **11**, 431–441.
- Mueller, S. G., Weiner, M. W., Thal, L. J., Petersen, R. C., Jack, C., Jagust, W., Trojanowski, J. Q., Toga, A. W., and Beckett, L. (2005). The Alzheimer’s disease Neuroimaging Initiative. *Neuroimaging Clinics of North America* **15**, 869–877.

- Mungas, D., Harvey, D., Reed, B., Jagust, W., DeCarli, C., Beckett, L., Mack, W., Kramer, J., Weiner, M., Schuff, N., et al. (2005). Longitudinal volumetric MRI change and rate of cognitive decline. *Neurology* **65**, 565–571.
- Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., and Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology* **37**, 323–329.
- Prague, M., Commenges, D., Gran, J., Ledergerber, B., Young, J., Furrer, H., and Thiebaut, R. (2017). Dynamic models for estimating the effect of HAART on CD4 in observational studies: application to the Aquitaine Cohort and the Swiss HIV Cohort Study. *Biometrics* **73**, 294–304.
- Proust-Lima, C., Amieva, H., and Jacqmin-Gadda, H. (2013). Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *British Journal of Mathematical and Statistical Psychology* **66**, 470–487.
- Proust-Lima, C., Jacqmin-Gadda, H., Taylor, J. M., Ganiayre, J., and Commenges, D. (2006). A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics* **62**, 1014–1024.
- Proust-Lima, C., Philipps, V., Dartigues, J. F., Bennett, D. A., Glymour, M. M., Jacqmin-Gadda, H., and Samieri, C. (2017). Are latent variable models preferable to composite score approaches when assessing risk factors of change? Evaluation of type-I error and statistical power in longitudinal cognitive studies. *Statistical Methods in Medical Research, in press*.
- Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science* **3**, 425–441.
- Reitz, C. and Mayeux, R. (2014). Alzheimer’s disease: epidemiology, diagnostic criteria, risk factors and biomarkers. *Biochemical pharmacology* **88**, 640–651.
- Robitaille, A., Muniz, G., Piccinin, A. M., Johansson, B., and Hofer, S. M. (2012). Multivariate longitudinal modeling of cognitive aging: Associations among change and variation in processing speed and visuospatial ability. *GeroPsych: The Journal of Gerontopsychology and Geriatric Psychiatry* **25**, 15.

- Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association* **100**, 322–331.
- Soetaert, K., Petzoldt, T., and Setzer, R. W. (2010). Solving differential equations in r: Package desolve. *Journal of Statistical Software* **33**, 1–25.
- Song, L., Kolar, M., and Xing, E. P. (2009). Time-varying dynamic bayesian networks. In *Advances in Neural Information Processing Systems*, pages 1732–1740.

Appendix A: Relationships between model components under different discretization steps

For computational reasons, the latent processes in the structural model are assumed to evolve in discrete time with a constant step δ which may vary depending on the application. In reality the processes evolve in continuous time and causal inference should be done in continuous time. We thus established the relationship between transition matrices obtained under different discretization steps: δ^* and a smaller discretization step δ with $\delta^* = \rho \times \delta$ and integer $\rho > 1$.

With a discretization step $\delta = \frac{\delta^*}{\rho}$, the second line of equation (1) can be rewritten:

$$\mathbf{\Lambda}_i(t + \delta) = \delta (\mathbf{X}_i(t + \delta)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta)\mathbf{v}_i) + (\mathbf{I}_D + \delta \mathbf{A}_{i,\delta}(t)) \mathbf{\Lambda}_i(t), \quad \forall t > 0, \rho > 1 \quad (17)$$

By recurrence, the network level at $t + \rho\delta$, that also correspond to $t + \delta^*$, can be expressed as a function of $\mathbf{\Lambda}_i(t + \delta^*)$ as follows:

$$\mathbf{\Lambda}_i(t + \delta^*) = \Psi_{i,\delta}(t, \rho, 0)\mathbf{\Lambda}_i(t) + \delta \sum_{s=1}^{\rho} \Psi_{i,\delta}(t, \rho, s) (\mathbf{X}_i(t + \delta s)\boldsymbol{\gamma} + \mathbf{Z}_i(t + \delta s)\mathbf{v}_i) \quad (18)$$

with function $\Psi_{i,\delta}$ defined in the relation (7).

Considering directly a discretization step of δ^* , the second line of equation (1) can also be rewritten

$$\mathbf{\Lambda}_i(t + \delta^*) = \delta^* (\mathbf{X}_i(t + \delta^*)\boldsymbol{\gamma}^* + \mathbf{Z}_i(t + \delta^*)\mathbf{v}_i^*) + (\mathbf{I}_D + \delta^* \mathbf{A}_{i,\delta^*}(t)) \mathbf{\Lambda}_i(t), \quad \forall t > 0 \quad \text{and} \quad \delta^* > 0. \quad (19)$$

Relationships between transition matrices under different steps Considering that the second parts of equations (18) and (19) are close enough although the model specifications are different, we obtain:

$$\mathbf{I}_D + \delta^* \mathbf{A}_{i,\delta^*}(t) \approx \Psi_{i,\delta}(t, \rho, 0) = \prod_{l=0}^{\rho-1} \left(\mathbf{I}_D + \frac{\delta^*}{\rho} \mathbf{A}_{i,\delta}(t + l \frac{\delta^*}{\rho}) \right) \quad (20)$$

Equation (20) can be rewritten to highlight the relationship between the causal structure \mathbf{A}_{i,δ^*} and $\mathbf{A}_{i,\delta}$:

$$\mathbf{A}_{i,\delta^*}(t) = \frac{1}{\delta^*} \left(\prod_{l=0}^{\rho-1} \left(\mathbf{I}_D + \frac{\delta^*}{\rho} \mathbf{A}_{i,\delta}(t + l \frac{\delta^*}{\rho}) \right) - \mathbf{I}_D \right) \quad (21)$$

When assuming that the causal structure $\mathbf{A}_{i,\delta}$ is constant in each interval $[t, t + \delta^*]$, $\forall t \in \tau$, the relationship (21) becomes :

$$\mathbf{A}_{i,\delta^*}(t) = \frac{1}{\delta^*} \left(\left(\mathbf{I}_D + \frac{\delta^*}{\rho} \mathbf{A}_{i,\delta}(t) \right)^\rho - \mathbf{I}_D \right) \quad (22)$$

Or equivalently,

$$\mathbf{A}_{i,\delta}(t) = \frac{\rho}{(\delta^*)^{(\rho+1)}} \left(\left(\frac{1}{\delta^*} \mathbf{I}_D + \mathbf{A}_{i,\delta^*}(t) \right)^{\frac{1}{\rho}} - \mathbf{I}_D \right) \quad (23)$$

Note that the limits of the right part of relation (22), when ρ tends to infinity provide the relationship between $\mathbf{A}_{i,\delta^*}(t)$ and its continuous time counterpart. Assuming that $\lim_{\rho \rightarrow \infty} \mathbf{A}_{i,\delta}(t) = \mathbf{A}_i(t)$ is finite, the relationship (22), become:

$$\mathbf{A}_{i,\delta^*}(t) \approx \frac{1}{\delta^*} (\exp(\delta^* \mathbf{A}_i(t)) - \mathbf{I}_D) \quad (24)$$

Relationships between trend parameters under different steps To relate the transition matrices under different steps, we considered second parts of equations (18) and (19) were close enough. We identified:

$$\begin{cases} \mathbf{X}_i(t + \delta^*) \boldsymbol{\gamma}^* \approx \frac{1}{\rho} \sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \mathbf{X}_i(t + \frac{\delta^*}{\rho} s) \boldsymbol{\gamma} \\ \mathbf{Z}_i(t + \delta^*) \mathbf{v}_i^* \approx \frac{1}{\rho} \sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \mathbf{Z}_i(t + \frac{\delta^*}{\rho} s) \mathbf{v}_i \end{cases} \quad (25)$$

In the particular case where \mathbf{X}_i and \mathbf{Z}_i reduce to intercepts,

$$\begin{cases} \boldsymbol{\gamma}^* \approx \frac{1}{\rho} \sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \boldsymbol{\gamma} \\ \mathbf{v}_i^* \approx \frac{1}{\rho} \sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \mathbf{v}_i \end{cases} \quad (26)$$

Or equivalently,

$$\begin{cases} \boldsymbol{\gamma} \approx \rho \left(\sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \right)^{-1} \boldsymbol{\gamma}^* \\ \mathbf{v}_i \approx \rho \left(\sum_{s=1}^{\rho} \boldsymbol{\Psi}_{i,\delta}(t, \rho, s) \right)^{-1} \mathbf{v}_i^* \end{cases} \quad (27)$$

Appendix B: Additional results in the simulations studies

Table 4: Results of the simulations (100 replicates of samples with 512 subjects) considering two latent processes, each one repeatedly measured by a marker, with linear trajectories and constant covariate-specific causal structure. ASE is the asymptotic standard error, ESE is the empirical standard error and CR is the coverage rate of the 95% confidence interval.

| | θ | without missing values (conv [†] =100%) | | | | with missing values * (conv [†] =99%) | | | |
|--------------|----------|--|------------------|------------------|-------|--|------------------|------------------|-------|
| | | $\hat{\theta}$ | ESE [‡] | ASE [‡] | CR(%) | $\hat{\theta}$ | ESE [‡] | ASE [‡] | CR(%) |
| β_1^1 | -0.268 | -0.267 | 0.068 | 0.073 | 97.0 | -0.269 | 0.070 | 0.075 | 97.9 |
| β_2^1 | -1.695 | -1.698 | 0.112 | 0.112 | 92.0 | -1.705 | 0.113 | 0.116 | 92.9 |
| β_1^2 | 0.057 | 0.055 | 0.075 | 0.078 | 96.0 | 0.060 | 0.079 | 0.082 | 95.9 |
| β_2^2 | -1.749 | -1.777 | 0.139 | 0.127 | 94.0 | -1.775 | 0.148 | 0.137 | 94.9 |
| γ_0^1 | 0.042 | 0.044 | 0.018 | 0.019 | 97.0 | 0.044 | 0.020 | 0.021 | 97.9 |
| γ_1^1 | -0.033 | -0.032 | 0.020 | 0.019 | 94.0 | -0.032 | 0.020 | 0.020 | 94.9 |
| γ_2^1 | -0.242 | -0.232 | 0.048 | 0.052 | 97.0 | -0.225 | 0.054 | 0.063 | 98.9 |
| γ_0^2 | -0.097 | -0.098 | 0.041 | 0.040 | 95.0 | -0.098 | 0.047 | 0.045 | 92.9 |
| γ_1^2 | -0.014 | -0.016 | 0.023 | 0.020 | 93.0 | -0.015 | 0.023 | 0.022 | 95.9 |
| γ_2^2 | -0.066 | -0.062 | 0.063 | 0.056 | 92.0 | -0.061 | 0.068 | 0.062 | 94.9 |
| L(3,1) | 0.188 | 0.181 | 0.027 | 0.031 | 97.0 | 0.177 | 0.033 | 0.038 | 97.9 |
| L(4,2) | 0.067 | 0.066 | 0.021 | 0.022 | 96.0 | 0.069 | 0.027 | 0.026 | 96.9 |
| L(3,3) | 0.149 | 0.145 | 0.013 | 0.014 | 95.0 | 0.143 | 0.014 | 0.016 | 93.9 |
| L(4,4) | 0.250 | 0.247 | 0.019 | 0.019 | 96.0 | 0.247 | 0.022 | 0.022 | 93.9 |
| a_{11}^0 | -0.230 | -0.223 | 0.031 | 0.035 | 98.0 | -0.219 | 0.037 | 0.043 | 96.9 |
| a_{11}^1 | 0.099 | 0.100 | 0.017 | 0.017 | 95.0 | 0.100 | 0.017 | 0.019 | 96.9 |
| a_{12}^0 | 0.118 | 0.120 | 0.024 | 0.025 | 92.0 | 0.121 | 0.027 | 0.027 | 93.9 |
| a_{12}^1 | -0.040 | -0.044 | 0.024 | 0.023 | 93.0 | -0.045 | 0.026 | 0.027 | 95.9 |
| a_{21}^0 | 0.095 | 0.095 | 0.022 | 0.023 | 95.0 | 0.094 | 0.024 | 0.025 | 95.9 |
| a_{21}^1 | 0.043 | 0.046 | 0.026 | 0.027 | 97.0 | 0.046 | 0.027 | 0.029 | 98.9 |
| a_{22}^0 | -0.399 | -0.400 | 0.043 | 0.042 | 97.0 | -0.401 | 0.054 | 0.050 | 90.9 |
| a_{22}^1 | 0.319 | 0.325 | 0.034 | 0.033 | 96.0 | 0.321 | 0.040 | 0.037 | 95.9 |
| η_{01} | 3.793 | 3.794 | 0.188 | 0.195 | 94.0 | 3.795 | 0.195 | 0.199 | 93.9 |
| η_{02} | 2.601 | 2.608 | 0.135 | 0.127 | 93.0 | 2.607 | 0.148 | 0.136 | 93.9 |
| η_{11} | 1.597 | 1.594 | 0.026 | 0.027 | 97.0 | 1.593 | 0.027 | 0.028 | 96.9 |
| η_{12} | 1.226 | 1.225 | 0.026 | 0.025 | 93.0 | 1.225 | 0.026 | 0.027 | 96.9 |
| σ_1 | 0.397 | 0.399 | 0.015 | 0.015 | 95.0 | 0.399 | 0.015 | 0.016 | 96.9 |
| σ_2 | 0.672 | 0.672 | 0.031 | 0.029 | 96.0 | 0.673 | 0.032 | 0.033 | 95.9 |

[†] Rate of convergence

* (15% missing occasions, 7% missing outcomes)

[‡] ASE is the asymptotic standard error, ESE is the empirical standard error and CR is the coverage rate of the 95% confidence interval.

Appendix C: Additional results in ADNI application

Table 5: Fixed effect estimates on baseline levels and changes of cerebral, cognitive and functional dimensions using a discretization step of 6 months.

| Parameter | Cerebral dim. | | | Cognitive dim. | | | Functional dim. | | |
|------------------------------|---------------|-------|---------|----------------|-------|---------|-----------------|-------|---------|
| | θ | s.e | p-value | θ | s.e | p-value | θ | s.e | p-value |
| <u>At baseline</u> | | | | | | | | | |
| Age (entry) | -0.592 | 0.065 | <0.001 | -0.232 | 0.049 | <0.001 | -0.051 | 0.062 | 0.412 |
| Male | -0.187 | 0.082 | 0.024 | -0.176 | 0.086 | 0.042 | - | - | - |
| High. Educ. | -0.256 | 0.108 | 0.018 | 0.493 | 0.114 | <0.001 | - | - | - |
| APOE | -0.287 | 0.086 | <0.001 | -0.347 | 0.090 | <0.001 | -0.071 | 0.095 | 0.453 |
| MCI | -0.756 | 0.101 | <0.001 | -1.614 | 0.116 | <0.001 | -1.609 | 0.125 | <0.001 |
| dAD | -1.504 | 0.128 | <0.001 | -3.000 | 0.159 | <0.001 | -3.567 | 0.186 | <0.001 |
| <u>On the rate of change</u> | | | | | | | | | |
| Intercept | -0.082 | 0.031 | 0.010 | 0.067 | 0.061 | 0.278 | 0.049 | 0.090 | 0.584 |
| MCI | -0.027 | 0.025 | 0.281 | -0.471 | 0.091 | <0.001 | -0.302 | 0.104 | 0.004 |
| dAD | -0.148 | 0.052 | 0.004 | -1.007 | 0.186 | <0.001 | -1.316 | 0.228 | < 0.001 |
| Male | -0.017 | 0.021 | 0.420 | -0.011 | 0.035 | 0.753 | - | - | - |
| High. Educ. | -0.009 | 0.027 | 0.712 | 0.243 | 0.068 | <0.001 | - | - | - |
| APOE | -0.045 | 0.009 | <0.001 | -0.142 | 0.040 | <0.001 | -0.120 | 0.066 | 0.067 |
| t | 0.009 | 0.009 | 0.291 | 0.001 | 0.014 | 0.977 | -0.038 | 0.015 | 0.013 |
| MCI \times t | 0.001 | 0.007 | 0.840 | -0.013 | 0.012 | 0.292 | -0.023 | 0.020 | 0.261 |
| dAD \times t | 0.037 | 0.013 | 0.004 | -0.062 | 0.026 | 0.016 | 0.001 | 0.034 | 0.968 |
| Male \times t | 0.009 | 0.006 | 0.134 | - | - | - | - | - | - |
| High. Educ. \times t | -0.008 | 0.008 | 0.291 | -0.011 | 0.013 | 0.372 | - | - | - |
| APOE \times t | - | - | - | - | - | - | -0.005 | 0.019 | 0.778 |

Table 6: Parameter estimates of the Cholesky decomposition of the variance-covariance matrix of overall random effects on baseline levels and changes of cerebral, cognitive and functional dimensions. Elements at 0 or 1 were fixed; standard errors of estimates are indicated in brackets below.

| | u^1 | u^2 | u^3 | v_1^1 | v_2^1 | v_1^2 | v_2^2 | v_1^3 | v_2^3 |
|---------|-------------------|-------------------|-------------------|-------------------|------------------|------------------|-------------------|-------------------|------------------|
| u^1 | 1 | | | | | | | | |
| u^2 | 0 | 1 | | | | | | | |
| u^3 | 0 | 0 | 1 | | | | | | |
| v_1^1 | 0.064 (0.016) | 0 | 0 | 0.085 (0.023) | | | | | |
| v_2^1 | -0.004 (0.003) | 0 | 0 | -0.012 (0.009) | 0.011 (0.004) | | | | |
| v_1^2 | 0 | 0.378 (0.063) | 0 | 0 | 0 | 0.142 (0.026) | | | |
| v_2^2 | 0 | -0.009 (0.005) | 0 | 0 | 0 | 0.014 (0.008) | -0.001 (0.080) | | |
| v_1^3 | 0 | 0 | 0.299 (0.064) | 0 | 0 | 0 | 0 | -0.501 (0.044) | |
| v_2^3 | 0 | 0 | -0.004 (0.009) | 0 | 0 | 0 | 0 | 0.087 (0.012) | 0.089 (0.009) |

Table 7: Parameter estimates of the measurement model. Parameters $(\sigma_1, \sigma_2, \sigma_3)$ represent the standard deviation of the measurement errors for the three composite scores. Other parameters given by series of six (η_{10} to η_{15} ; η_{20} to η_{25} and η_{30} to η_{34}) correspond to the parameters of the quadratic I-splines link function used to transform the anatomic composite score, the cognitive composite score and the functional composite score, respectively.

| score | Parameter | Estimate | SE | p-value |
|------------------------|-------------|----------|-------|---------|
| cerebral anatomy score | η_{10} | -5.611 | 0.249 | <0.001 |
| | η_{11} | 0.527 | 0.115 | <0.001 |
| | η_{12} | 0.263 | 0.097 | 0.006 |
| | η_{13} | 0.846 | 0.046 | <0.001 |
| | η_{14} | 0.671 | 0.074 | <0.001 |
| | η_{15} | 0.558 | 0.208 | 0.007 |
| | σ_1 | 0.236 | 0.008 | <0.001 |
| cognitive score | η_{20} | -6.429 | 0.310 | <0.001 |
| | η_{21} | -0.122 | 0.285 | 0.668 |
| | η_{22} | 0.579 | 0.107 | <0.001 |
| | η_{23} | 1.436 | 0.042 | <0.001 |
| | η_{24} | 0.461 | 0.101 | <0.001 |
| | η_{25} | 0.884 | 0.136 | <0.001 |
| | σ_2 | 0.394 | 0.015 | <0.001 |
| functional score | η_{30} | -7.332 | 0.299 | <0.001 |
| | η_{31} | 0.953 | 0.082 | <0.001 |
| | η_{32} | 0.272 | 0.135 | 0.045 |
| | η_{33} | 0.528 | 0.064 | <0.001 |
| | η_{34} | 0.631 | 0.041 | <0.001 |
| | σ_3 | 0.581 | 0.025 | <0.001 |

Table 8: Estimates of the transition matrix intensities on ADNI data when considering three discretization steps: 6 months (AIC=1386.68), 3 months (AIC=1370.45) and 2 months (AIC=1367.21).

| | $\delta = 6$ months | | $\delta = 3$ months | | $\delta = 2$ months | |
|-----------------|---------------------|-------|---------------------|-------|---------------------|-------|
| | Estimate | SE | Estimate | SE | Estimate | SE |
| α_{11}^0 | -0.054 | 0.014 | -0.054 | 0.015 | -0.054 | 0.015 |
| α_{11}^1 | 0.008 | 0.009 | 0.008 | 0.009 | 0.008 | 0.009 |
| α_{11}^2 | 0.001 | 0.013 | 0.001 | 0.013 | 0.001 | 0.013 |
| α_{12}^0 | 0.010 | 0.008 | 0.011 | 0.008 | 0.011 | 0.008 |
| α_{12}^1 | 0.022 | 0.010 | 0.022 | 0.010 | 0.022 | 0.010 |
| α_{12}^2 | 0.025 | 0.014 | 0.025 | 0.015 | 0.025 | 0.015 |
| α_{13}^0 | 0.010 | 0.017 | 0.008 | 0.017 | 0.008 | 0.017 |
| α_{13}^1 | -0.012 | 0.017 | -0.011 | 0.018 | -0.010 | 0.018 |
| α_{13}^2 | -0.011 | 0.019 | -0.010 | 0.020 | -0.009 | 0.020 |
| α_{21}^0 | 0.077 | 0.023 | 0.083 | 0.026 | 0.084 | 0.027 |
| α_{21}^1 | 0.018 | 0.026 | 0.019 | 0.029 | 0.019 | 0.029 |
| α_{21}^2 | -0.009 | 0.034 | -0.010 | 0.037 | -0.011 | 0.039 |
| α_{22}^0 | -0.396 | 0.074 | -0.423 | 0.088 | -0.432 | 0.093 |
| α_{22}^1 | 0.081 | 0.030 | 0.088 | 0.033 | 0.090 | 0.034 |
| α_{22}^2 | 0.026 | 0.042 | 0.029 | 0.045 | 0.029 | 0.047 |
| α_{23}^0 | 0.146 | 0.054 | 0.156 | 0.062 | 0.159 | 0.064 |
| α_{23}^1 | -0.068 | 0.054 | -0.073 | 0.060 | -0.074 | 0.062 |
| α_{23}^2 | -0.078 | 0.059 | -0.080 | 0.066 | -0.079 | 0.068 |
| α_{31}^0 | 0.077 | 0.035 | 0.103 | 0.048 | 0.109 | 0.050 |
| α_{31}^1 | 0.025 | 0.043 | 0.017 | 0.056 | 0.016 | 0.058 |
| α_{31}^2 | -0.001 | 0.052 | -0.010 | 0.068 | -0.012 | 0.071 |
| α_{32}^0 | 0.132 | 0.042 | 0.177 | 0.062 | 0.185 | 0.063 |
| α_{32}^1 | 0.018 | 0.048 | 0.000 | 0.066 | -0.004 | 0.066 |
| α_{32}^2 | 0.012 | 0.061 | -0.006 | 0.079 | -0.009 | 0.082 |
| α_{33}^0 | -0.892 | 0.108 | -1.195 | 0.207 | -1.237 | 0.219 |
| α_{33}^1 | 0.569 | 0.078 | 0.770 | 0.142 | 0.798 | 0.149 |
| α_{33}^2 | 0.478 | 0.082 | 0.667 | 0.141 | 0.692 | 0.148 |

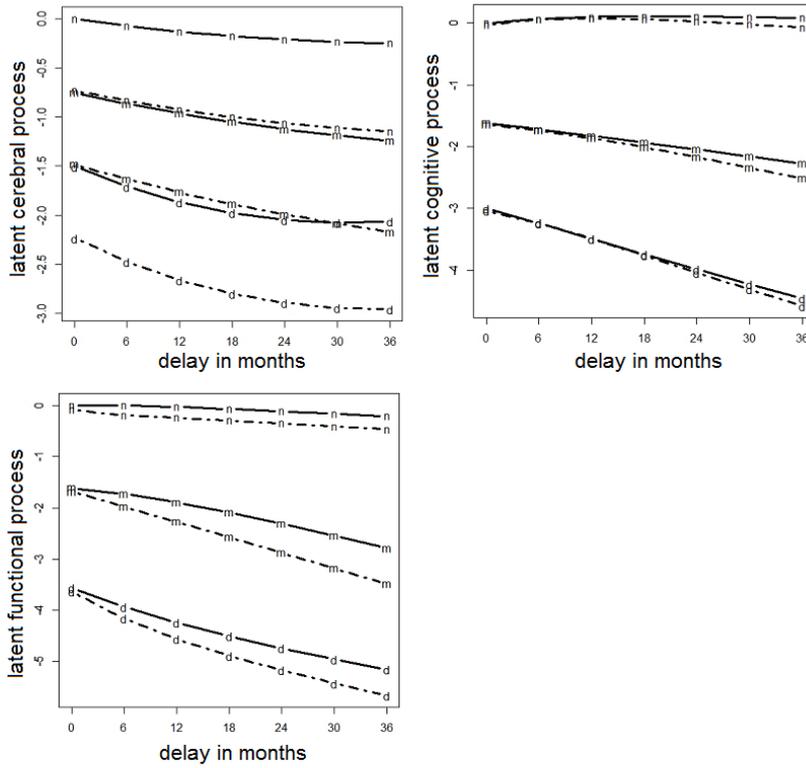


Figure 4: Predicted trajectories of cerebral, cognitive and functional dimensions at healthy (CN), MCI, and dAD clinical stages of Alzheimer’s disease for a woman non-carrier of APOE $\epsilon 4$ allele with lower educational level (plain lines) and for a man carrier of APOE $\epsilon 4$ allele with higher educational level (dashed lines). “n”, “m” and “d” indicate trajectories at healthy, MCI and dAD stages, respectively.

6.2 Complément de résultats de l'application 1

TABLE A 6.1 – Estimations des paramètres non contraints de la décomposition de Cholesky (\mathbf{L}) de la matrice de variance-covariance $\mathbf{B} = \mathbf{L}^\top \mathbf{L}$ des effets aléatoires $(u_{i1}, u_{i2}, u_{i3}, v_{i1}^1, v_{i1}^2, v_{i2}^1, v_{i2}^2, v_{i3}^1, v_{i3}^2)^\top$ du modèle multivarié de la dimension anatomique mesurée par le composite score de 10 régions du cerveau (**Anat10**), de la dimension cognitive mesurée par le composite score de 19 scores psychométriques standardisés (**Cog19**) et la dimension fonctionnelle mesurée par la somme standardisée de 30 items de la FAQ (**Fct**); les taux de variation des dimensions sont supposés linéaires au cours du temps (modèle de l'application 1 de la section 5.2.1.1).

| Paramètre | Estimation | ET | p-valeur |
|-------------------|------------|-------|----------|
| $L_{u_2 u_1}$ | 0.357 | 0.045 | < 0.001 |
| $L_{u_3 u_1}$ | 0.176 | 0.048 | < 0.001 |
| $L_{v_1^1 u_1}$ | 0.109 | 0.029 | < 0.001 |
| $L_{v_1^2 u_1}$ | -0.013 | 0.011 | 0.222 |
| $L_{u_3 u_2}$ | 0.286 | 0.050 | < 0.001 |
| $L_{v_2^1 u_2}$ | 0.723 | 0.153 | < 0.001 |
| $L_{v_2^2 u_2}$ | -0.019 | 0.021 | 0.357 |
| $L_{v_3^1 u_3}$ | 0.123 | 0.159 | 0.439 |
| $L_{v_3^2 u_3}$ | 0.029 | 0.036 | 0.407 |
| $L_{v_1^1 v_1^1}$ | 0.156 | 0.037 | < 0.001 |
| $L_{v_1^2 v_1^1}$ | -0.033 | 0.032 | 0.293 |
| $L_{v_1^2 v_1^2}$ | 0.034 | 0.021 | 0.102 |
| $L_{v_2^1 v_2^1}$ | 0.307 | 0.089 | < 0.001 |
| $L_{v_2^2 v_2^1}$ | 0.047 | 0.086 | 0.586 |
| $L_{v_2^2 v_2^2}$ | -0.050 | 0.182 | 0.783 |
| $L_{v_3^1 v_3^1}$ | 0.855 | 0.083 | < 0.001 |
| $L_{v_3^2 v_3^1}$ | -0.358 | 0.040 | < 0.001 |
| $L_{v_3^2 v_3^2}$ | 0.209 | 0.034 | < 0.001 |

TABLE A 6.2 – Estimations des écart-types des erreurs de mesure σ_{Anat10} du marqueur **Anat10** (composite score de 10 régions du cerveau qui mesure la dimension anatomique), σ_{Cog19} du marqueur **Cog19** (composite score de 19 scores psychométriques standardisés qui mesure la dimension cognitive), σ_{Fct} du marqueur **Fct** (somme standardisée de 30 items de la FAQ qui mesure la dimension fonctionnelle) et des paramètres des fonctions de lien par I-splines intégrées quadratiques avec 4 noeuds pour les marqueurs **Anat10**, **Cog19** et 3 noeuds pour le marqueur **Fct** (cf. modèle principal de l'application 1 de la section 5.2.1.1).

| Paramètre | Estimation. | ET | p-valeur |
|--------------------------|-------------|-------|----------|
| σ_{Anat10} | 0.237 | 0.008 | < 0.001 |
| σ_{Cog19} | -0.418 | 0.017 | < 0.001 |
| σ_{Fct} | 0.632 | 0.029 | < 0.001 |
| Anat10.4-quant-2.1 | -5.625 | 0.248 | < 0.001 |
| Anat10.4-quant-2.2 | 0.518 | 0.115 | < 0.001 |
| Anat10.4-quant-2.3 | 0.274 | 0.096 | 0.004 |
| Anat10.4-quant-2.4 | 0.852 | 0.046 | < 0.001 |
| Anat10.4-quant-2.5 | 0.670 | 0.075 | < 0.001 |
| Anat10.4-quant-2.6 | 0.556 | 0.209 | 0.008 |
| Cog19.4-quant-2.1 | -6.790 | 0.326 | < 0.001 |
| Cog19.4-quant-2.2 | -0.044 | 0.274 | 0.872 |
| Cog19.4-quant-2.3 | 0.581 | 0.111 | < 0.001 |
| Cog19.4-quant-2.4 | 1.510 | 0.042 | < 0.001 |
| Cog19.4-quant-2.5 | 0.475 | 0.105 | < 0.001 |
| Cog19.4-quant-2.6 | 0.981 | 0.132 | < 0.001 |
| Fct.3-quant-2.1 | -7.827 | 0.333 | < 0.001 |
| Fct.3-quant-2.2 | 1.140 | 0.076 | < 0.001 |
| Fct.3-quant-2.3 | -0.289 | 0.248 | 0.244 |
| Fct.3-quant-2.4 | 0.831 | 0.056 | < 0.001 |
| Fct.3-quant-2.5 | 0.605 | 0.043 | < 0.001 |

Modélisation dynamique des sphères anatomique, cognitive et fonctionnelle dans la maladie d'Alzheimer : une approche par processus latents

En santé publique, l'étude de la progression d'une maladie chronique et de son mécanisme peut nécessiter la modélisation conjointe de plusieurs marqueurs longitudinaux et leur structure de dépendances. Des approches de modélisation existent dans la littérature pour répondre partiellement aux objectifs de modélisation. Ces approches deviennent rapidement coûteuses et difficiles à utiliser dans certaines maladies complexes à caractère latent, dynamique et multidimensionnel, comme la maladie d'Alzheimer. Dans un tel contexte, l'objectif de cette thèse était de proposer une méthodologie innovante pour modéliser la dynamique de plusieurs processus latents et leurs influences temporelles à des fins d'interprétations causales à partir d'observations répétées de marqueurs continus Gaussiens et non Gaussiens. L'approche proposée, centrée sur des processus latents, définit un modèle structurel pour la trajectoire des processus latents et un modèle d'observation pour lier les marqueurs longitudinaux aux processus qu'ils mesurent. Dans le modèle structurel, défini en temps discret, le niveau initial et le taux de variation des processus spécifiques aux individus sont modélisés par des modèles linéaires à effets mixtes. Le modèle du taux de variation comporte une composante auto-régressive d'ordre 1 qui permet de modéliser l'effet d'un processus sur un autre processus en tenant explicitement compte du temps. Le modèle structurel, tel que défini, bénéficie des mêmes interprétations causales que les modèles à équations différentielles (ODE) de l'approche mécaniste de la causalité tout en évitant les problèmes numériques de ces derniers. Le modèle d'observation utilise des fonctions de lien paramétrées pour que le modèle puisse être appliqué à des marqueurs longitudinaux possiblement non Gaussiens. La méthodologie a été validée par des études de simulations. Cette approche, appliquée à la maladie d'Alzheimer a permis de décrire conjointement la dynamique de l'atrophie de l'hippocampe, du déclin de la mémoire épisodique, du déclin de la fluence verbale et de la perte d'autonomie ainsi que les influences temporelles entre ces dimensions dans plusieurs phases de la maladie à partir des données ADNI.

Mots clés : données répétées ; processus latents ; causalité ; modèles mixtes ; équations de différence ; maladie d'Alzheimer.

Dynamic modeling of anatomic cognitive and functional spheres in Alzheimer's disease : a latent process approach

In public health, the study of the progression of a chronic disease and its mechanisms may require the joint modeling of several longitudinal markers and their dependence structure. Modeling approaches exist in the literature to partially address these modeling objectives. But these approaches become rapidly numerically expensive and difficult to use in some complex diseases involving latent, dynamic and multidimensional aspects, such as in Alzheimer's disease. The aim of this thesis was to propose an innovative methodology for modeling the dynamics of several latent processes and their temporal influences for the purpose of causal interpretations, from repeated observations of continuous Gaussian and non Gaussian markers. The proposed latent process approach defines a structural model in discrete time for the latent processes trajectories and an observation model to relate longitudinal markers to the process they measure. In the structural model, the initial level and the rate of change of individual-specific processes are modeled by mixed-effect linear models. The rate of change model has a first order auto-regressive component that can model the effect of a process on another process by explicitly accounting for time. The structural model as defined benefits from the same causal interpretations as the models with differential equations (ODE) of the mechanistic approach of the causality while avoiding major numerical problems. The observation model uses parameterized link functions to handle possibly non-Gaussian continuous markers. The consistency of the ML estimators and the accuracy of the inference of the influence structures between the latent processes have been validated by simulation studies. This approach, applied to Alzheimer's disease, allowed to jointly describe the dynamics of hippocampus atrophy, the decline of episodic memory, the decline of verbal fluency, and loss of autonomy as well as the temporal influences between these dimensions in several stages of Alzheimer's dementia from the data of the ADNI initiative.

Key words : longitudinal data ; latent processes ; causality ; mixed models ; difference equations ; Alzheimer's disease.

Spécialité : Santé publique – option : Biostatistique

Laboratoire : Unité Inserm U1219 - Université de Bordeaux - Isped
146 rue Léo Saignat, 33076 Bordeaux Cedex, France